

Universität
Rostock



Traditio et Innovatio

**Underwater Image Restoration:
Super-resolution and Deblurring via Sparse Representation and
Denoising by Means of Marine Snow Removal**

Dissertation

zur

Erlangung des akademischen Grades

Doktor-Ingenieur (Dr.-Ing)

der Fakultät für Informatik und Elektrotechnik

der Universität Rostock

vorgelegt von

Fahimeh Farhadifard

geb. am 05.11.1985 in Mashhad/Iran

aus Rostock

Rostock, den 27. Oktober 2017

Examiners:

Prof. Dr.-Ing. Bodo Urban
Universität Rostock, Fraunhofer IGD, Germany

Prof. Dr.-Ing. Uwe Freiherr von Lukas
Universität Rostock, Fraunhofer IGD, Germany

Prof. Dr.-Ing. Reinhard Koch
Christian-Albrechts-Universität zu Kiel, Germany

Date of Defense:

20. April 2018

Keywords:

Image restoration, Image Super-resolution, Image Deblurring, Dictionary Learning, Sparse Representation, Background Subtraction, Image Priors

Classification (ACM CCS2012):

Computing methodologies → Computer graphics → Image manipulation → Image Processing

Abstract

Underwater imaging has been widely used as a tool in many fields, such as marine industry, deep-sea mining, aquaculture and water assessment. However, a major issue is the quality of the resulting images and videos. Due to the light's interaction with water and its constituents, the acquired underwater images and videos often suffer from a significant amount of scatter (blur and haze) and noise. Furthermore, since data transmission from the equipment mounted under water to the station above water is still a challenge, usually a compressed and low-resolution version of the data is transferred.

In the light of these issues, this thesis considers the problems of low-resolution, blurred and noisy underwater images and proposes several approaches to improve the quality of such images/video frames. This is undertaken through two main contributions.

The first major contribution of this work is the super-resolution and deblurring of single underwater images. This is done by using a set of compact high and low-resolution cluster dictionaries where sparse representation is used as the regularizer. Since such an approach inevitably calls for a model selection criterion in both learning and reconstruction stages, a scale-invariance model is proposed to properly establish the link between the low and high-resolution feature spaces.

The subject of the second major contribution is image denoising. Besides additive noises such as sensor noise, the visibility in underwater images is reduced by the presence of suspended particles in water. This represents an unwanted signal, which is also disruptive for advanced computer vision tasks, such as segmentation. Since this phenomenon is a real signal and part of the scene, two-fold approaches consisting of first detection and then removal of such particles, are proposed. To avoid the uncertainty introduced by using local information for restoration, some global priors of the scene are learned, which are then used to estimate the parts of the scene that are covered by the particles. For this, a Gaussian-based background subtraction approach is proposed to obtain static features of the scene. These are used as training data for learning the priors.

Quantitative and qualitative experiments conducted over real and simulated underwater images and video frames validate the success of the proposed approaches at improving the image resolution and deblurring image features significantly as well as detecting and removing marine particles, while the object edges are preserved.

Kurzfassung

Unterwasseraufnahmen sind ein weitverbreitetes Hilfsmittel in vielen Anwendungsfeldern, wie zum Beispiel in der Schiffsindustrie, dem Meeresbodenbergbau, der Aquakultur und der Wasseranalyse. Die Qualität der resultierenden Bilder und Videos ist jedoch für viele gewünschte Auswertungen nicht ausreichend, da die Interaktion des Lichts mit dem Wasser und dessen Bestandteilen zu deutlicher Unschärfe und Trübung führt. Eine weitere Herausforderung stellt die Datenübertragung von Unterwasserinstallationen zu einer Station über Wasser dar, weil große Datenmengen für Echtzeit-Anwendungen komprimiert und niedrig aufgelöst übertragen werden müssen.

In dieser Dissertation werden verschiedenen Verfahren für die Qualitätsverbesserung von niedrig aufgelösten, unscharfen und verrauschten Unterwasserfotos sowie Videoeinzeln Bildern vorgestellt.

Ein besonderes Ergebnis dieser Arbeit besteht in der Bildverbesserung durch Superauflösung von einzelnen Unterwasserbildern und in der Reduktion der Unschärfe. Dies wird durch den Einsatz einer Reihe von kompakten hoch und niedrig aufgelösten, angelernten "Cluster Dictionaries" erreicht, wobei eine "Sparse Representation" als Regulierungsfunktion dient. Da ein solches Vorgehen zwangsläufig ein Kriterium für die Auswahl eines Modells sowohl für die Lern- als auch für die Rekonstruktionsphasen verlangt, wird ein skalierungsinvariantes Modell vorgeschlagen, das die Verbindung zwischen den hoch und niedrig aufgelösten Merkmalsräumen herstellt.

Das zweite hervorzuhebende Ergebnis besteht in der Rauschunterdrückung. Neben dem Sensorrauschen bewirken auch Schwebeteilchen im Wasser eine Verschlechterung der Sichtverhältnisse und beeinträchtigen damit die Qualität von Unterwasseraufnahmen. Schwebeteilchen erzeugen unerwünschte Signale, die sich in nachgelagerten Computer-Vision-Verfahren störend auswirken, wie zum Beispiel bei Segmentierungen. Für das Entfernen von Schwebeteilchen aus dem Bildmaterial wurde ein zweistufiges Verfahren entwickelt, bestehend aus der Detektion und nachfolgend der Entfernung und Rekonstruktion. Für die Abschätzung einer adäquaten Restauration der durch Schwebeteilchen verdeckten Regionen der Szene werden globale Merkmale angelernt, die im Ergebnis eine höhere Treffsicherheit als Verfahren mit Auswertungen lokaler Informationen erreichen. Es kommt ein auf Wahrscheinlichkeitsfunktionen basierendes Hintergrundsubtraktionsverfahren zum Einsatz, so dass statische Merkmale der Szene separiert werden können, die als Trainingsdaten für das Anlernen der globalen Merkmale genutzt werden können.

Quantitative und qualitative Experimente mit realen und simulierten Unterwasserfotos und Videoeinzeln Bildern bestätigen den Erfolg der vorgestellten Verfahren für die Verbesserung der Bildauflösung, zum Hervorheben von Bildmerkmalen und für die Detektion und Beseitigung von Schwebeteilchen einschließlich Restaurierung.

Acknowledgments

I would like to express my sincere gratitude to my advisor Prof. Dr. Ing. Bodo Urban for continues support of my Ph.D study and related research, for his patience, motivation and immense knowledge. Further, his guidance throughout my study which allowed me to develop my skills.

I would like to thank Prof. Dr. Ing. Uwe Freiherr von Lukas for his intention and openness which gave me enormous encouragement in every stage of the preparation of this thesis.

My sincere thanks also goes to Prof. Dr. Ing. Reinhard Koch for his insightful comments and encouragement which incentivized me to widen my research from various perspectives.

I thank my fellow colleagues at Fraunhofer IGD for the stimulating discussions, for the late evenings we were working together before deadlines, and for all the fun we had in the last three years.

Last but not least, I would like to express my love, gratitude and respect towards my family for supporting and encouraging me throughout this thesis and my life in general.

Contents

Abstract	III
Acknowledgments	V
List of Figures	IX
List of Tables	XII
Abbreviations and Symbols	XV
1 Introduction	1
1.1 Thesis Contributions	3
1.2 Outline of the Dissertation	5
2 Underwater Imaging and Degradations	7
2.1 Introduction	7
2.2 Underwater Effects	7
2.3 Underwater Image Formation	9
2.4 Jaffe-McGlamery Model	10
2.4.1 The Direct and Forward-Scattered Components	11
2.4.2 The Backscatter Component	12
3 Literature Review	15
3.1 Introduction	15
3.2 Underwater Image Restoration	16
3.2.1 Image Restoration based on Point Spread Function Estimation	16
3.2.2 Image Restoration Using Statistical Prior	22
3.2.3 Other Approaches	28
3.3 Underwater Image Enhancement	30
3.3.1 Histogram and Contrast Stretching-based Approaches	31
3.3.2 Filter-based Enhancement Approaches	35
3.3.3 Other Approaches	38
3.4 Underwater Image Denoising	40
3.5 Summary	44
3.6 Conclusion	47

4	Sparse Representation: Definition and Applications	49
4.1	Introduction	49
4.1.1	Problem Formulation	50
4.1.2	Matching Pursuit Family (MPF)	51
4.1.3	Convex Relaxation Algorithms	53
4.1.4	Choice of Dictionary	54
4.1.5	Application of Sparse Representation using Dictionary Learning	57
4.2	Single Image Super-Resolution (SISR)	61
4.2.1	K-SVD Approach	62
4.2.2	Orthogonal Matching Pursuit (OMP)	63
5	Underwater Single Image Super-Resolution and Deblurring via Sparse Representation	65
5.1	Introduction	65
5.2	Underwater Blur and Its Degradation Model	67
5.3	The Blur Measure (BM)	69
5.4	Clustering and Sparse Model Selection with BM	72
5.4.1	Training Phase	72
5.4.2	Reconstruction Phase	74
5.5	Experimental Results and Discussion	75
5.5.1	Approximate Scale-Invariance and Generality of the Proposed Blur Measure	75
5.5.2	Effect of the Patch Size and the Dictionary Redundancy on the Representation Quality	78
5.5.3	Implementation Details and Evaluations	80
5.6	Conclusion	91
6	Denoising: Marine Snow Detection and Removal	95
6.1	Introduction	95
6.2	Marine Snow	96
6.2.1	Marine Snow as a Source of Noise and its Characteristics	97
6.3	Related Works	101
6.4	Single Image Marine Snow Removal via Median ² Filtering	103
6.4.1	Coarse Detection	104
6.4.2	Fine Detection and Filtering	105
6.4.3	Simulation Results and Discussions	106
6.5	Marine Snow Removal from Video Frames with a Static Camera	114
6.5.1	Background Modeling	115
6.5.2	Statistical Priors Modeling using the Background Model	118
6.5.3	Inpainting	121
6.5.4	Marine Snow Removal	121
6.5.5	Simulation Results and Discussions	122
6.6	Conclusion	126
7	Combined Super-Resolution, Deblurring and Marine Snow Removal	129

8 Conclusion and Future Work **133**

8.1 Conclusion 133

8.2 Future Work 135

References **137**

Own Publications **151**

Thesis Statements **153**

List of Figures

2.1	Illustration of light interaction with water.	8
2.2	Examples of underwater image degradations.	9
2.3	Coordinate system of Jaffe-McGlamery’s model.	11
3.1	Performance of the [TS09] approach.	17
3.2	Performance of the [TO06] approach.	18
3.3	Performance of the [WL13] approach.	19
3.4	Performance of the [Hou+07b] approach.	20
3.5	Light-path components of the [SB14] approach (a)	21
3.6	Light-path components of the [SB14] approach (b)	21
3.7	Performance of the [SB14] approach.	22
3.8	Performance of the [CME10] approach.	24
3.9	Performance of the [CC12] approach.	25
3.10	Performance of the [PC17] approach.	28
3.11	Performance of the [TD05] approach.	29
3.12	Performance of the [Lu+17] approach.	30
3.13	Performance of the [Iqb+10] approach.	32
3.14	Performance of the [Hit+13] approach.	33
3.15	Performance of the [BG12] approach.	34
3.16	Performance of the [Baz+06] approach.	37
3.17	Performance of the [PCC09] approach.	38
3.18	Performance of the [Anc+12] approach.	39
3.19	Performance of the [SSS13] approach.	41
3.20	Performance of the [Liu+15a] approach.	42
3.21	Performance of the [Ban+14] approach.	43
5.1	An example of artifacts caused using a single dictionary pair.	66
5.2	Examples of different types of blur underwater.	67
5.3	Visualization of two learned dictionaries using sharp and blurred image patches.	69
5.4	A toy example to clarify the logic behind the proposed classifier.	71
5.5	Error in representation as a blur measure.	72
5.6	Representation of error values vs. blur level.	76
5.7	Histogram of the proposed blur measure for high and low-resolution image patches.	77
5.8	Average PSNR vs. dictionary redundancy.	79
5.9	Examples of the training images.	80
5.10	Average PSNR vs. number of clusters.	81
5.11	The learned HR dictionaries corresponding to 6 different clusters.	83

List of Figures

5.12	Visual comparison of the <i>coral reef 1</i> image using different algorithms.	85
5.13	Visual comparison of the <i>coral reef 2</i> image using different algorithms.	87
5.14	Visual comparison of the <i>submarine</i> image using different algorithms.	89
5.15	Visual comparison of the <i>statue</i> image using different algorithms.	90
5.16	Examples of estimated distance map using BM.	92
6.1	Examples of images distorted by marine snow.	98
6.2	Performance of different image processing algorithms in the presence of marine snow.	99
6.3	Illustration of marine snow characteristics.	100
6.4	Example of images with marine snow which an artificial light is used.	101
6.5	Visualization of RGB pixel values in the Euclidean space.	105
6.6	Visualization of double checking process (a).	106
6.7	Visualization of double checking process (b).	107
6.8	Visual comparison of the proposed Median ² approach with the literature (a) . .	108
6.9	Visual comparison of the proposed Median ² approach with the literature (b) . .	109
6.10	Visual comparison of the proposed Median ² approach with the literature (c) . .	111
6.11	Examples of simulated images with marine snow using the [BG12] approach. .	112
6.12	Examples of the problems that may occur using Median ² approach.	113
6.13	Comparison of different update schemes for the background modelling.	116
6.14	Examples of the background modelling.	122
6.15	Performance of the proposed video marine snow removal (a).	124
6.16	Performance of the proposed video marine snow removal (a).	125
7.1	Performance of the proposed methods for simulated data in a pipeline.	131
7.2	Performance of the proposed methods for real data in a pipeline.	132

List of Tables

3.1	Summary of UW Image Restoration Algorithms.	45
3.2	Summary of UW Image Enhancement Algorithms.	46
5.1	MSE performance of the dictionary pairs at restoring HR image patches in different clusters.	82
5.2	Quantitative comparisons using UISM, UIConM and entropy.	91
6.1	Quantitative Comparison of the proposed Median ² approach with the literature.	113
6.2	Quantitative evaluation using PSNR and MSE values.	123
7.1	Quantitative evaluation of the proposed methods via PSNR, MSE, UISM and Entropy.	130

Abbreviations and Symbols

Abbreviations

ACE	Automatic Color Equalization
AUV	Autonomous Underwater Vehicle
BM	Background Modeling
BM	Blur Measure
BM	Blur Measure
BP	Basis Pursuit
BRISQUE	Blind/Reference Image Spatial Quality Evaluator
CLAHE	Contrast Limited Adaptive Histogram Equalization
DL	Dictionary Learning
DPV	Discarded Pixel Values
EME	Enhancement Measure Estimation
FOCUSS	Focal Under-determined System Solver
FoE	Field of Experts
GSM	Gaussian Switch Model
HF	High-Frequency
HR	High-Resolution
IOP	Inherent Optical Properties
ISD	Image Signature Dictionary
LR	Low-Resolution
MAP	Maximum a posteriori Probability
MOD	Method of Optimal Directions
MOG	Mixture of Gaussian
MP	Matching Pursuit
MRF	Markov Random Field
MSE	Mean Square Error
MSR	Multi-Scale Retinex
MS	Marine Snow
MTF	Modulation Transfer Function
Med	Median
NP	Non-deterministic Polynomial-time
ODL	Online Dictionary Learning
OMP	Orthogonal Matching Pursuit
OOMP	Optimized Orthogonal Matching Pursuit
PCA	Principle Component Analysis
PSF	Point Spread Function

Abbreviations and Symbols

PSNR	Peak Signal to Noise Ratio
PoE	Product of Experts
ROV	Remotely Operated Vehicle
RPV	Remaining Pixel Values
SGD	Stochastic Gradient Descent
SISR	Single Image Super-Resolution
SR	Super-Resolution
SSIM	Structural Similarity
SVD	Singular Value Decomposition
UCIQE	Underwater Color Image Quality Evaluation
UICM	Underwater Image Colorfulness Measure
UIConM	Underwater Image Contrast Measure
UIQM	Underwater Image Quality Measure
UISM	Underwater Image Sharpness Measure
UW	Underwater
WCID	Wavelength Compensation and Image Dehazing
WGSA	Weighted Gray-scale Angles
logAMEE	Logarithmic Average Michelson Contrast Measure by Entropy

Symbols

A	Matrix of sparse coefficients
\vec{d}	Dictionary atom
D	Matrix, Dictionary
D_h	Matrix, High-resolution Dictionary
D_l	Matrix, Low-resolution Dictionary
E	Error
$E(\cdot)$	Irradiance
$f^{(\cdot)}$	Flag
$f(x, y)$	Point Spread Function
$I(x, y)$	Image
$J(x, y)$	Hazy Image
$L_{e/o/i}$	Emitted/outgoing/ingoing Radiance
$n(x, y)$	Noise
N	Noisy image
p	Pixel
P	Image Patch
\vec{r}	Residual vector
$R_{c/s}$	Object-camera Distance/ Object-source Distance
W	Weight
x	Vectorized high-resolution image patch
x_a	Vector of available pixels in a patch
x_m	Vector of missing pixels in a patch
X	High-resolution image
y	Vectorized low-resolution image patch
Y	Low-resolution image
z	Vectorized image patch
α	Sparse coefficient vector
$\ \cdot\ _F$	Frobenius Norm
$\ \cdot\ _p$	Norm p , $p \in \mathbb{R}$
ω	Spatial Frequency
Ω	Matrix Ω
\dagger	Pseudo inverse

1 Introduction

Digital imaging has been established in a broad area of challenging topics, such as in surveillance tasks, industrial quality assurance, inspection applications and exploration. This means that, digital images are encountered almost everywhere, from the internet or mobile devices to hospitals or satellites. Thus, the ability to process and enhance them is of paramount importance, and this is reflected by the wide range of advanced techniques for image processing that lies behind our everyday technology. Examples of such technologies can be seen in modern digital cinema and television, medical image analysis programs, or consumer software such as Photoshop or Instagram. From a more technical point of view, pre-processing input images may represent an essential step in order to improve advanced computer vision tasks.

Although the area of mathematical image processing and enhancement is quite mature it is nevertheless, a remarkably rich and lively field of research. In this sense, there are still strong limitations in the current state-of-the-art to properly process and enhance the quality of images whenever they are acquired in a non-controlled or hostile environment. It is, therefore, necessary to develop robust and efficient tools, capable of capturing the physical characteristics of each particular problem. An example of such a situation is underwater (UW) images.

The interest in UW videography has significantly increased since the invention of the first *amphibious* camera in 1960. In addition to divers, autonomous underwater vehicles (AUV) and remotely operated vehicles (ROV) are usually employed to explore the deep sea. These robots can reach depths where divers cannot operate safely and effectively. Despite the technological advances in the equipment, the quality of UW images and videos remain inferior to images shot in the air. Essentially, this is because of the limitations imposed by the inherent properties of water medium. Inherent optical properties (IOPs) are those properties that depend only upon the medium. The two fundamental IOPs usually employed in ocean optics are the absorption coefficient and the volume scattering function, which are respectively the spectral absorbance and scattering per unit distance in the medium. These effects contribute to serve UW image degradations in forms of:

Color cast, this is due to the fact that light loses its energy when traveling deeper through water. Thus, colors are reduced one by one, based on their wavelength.

Haze, light is scattered to the camera not only from the object of interest but also via water molecules and floating particles in the water. This signal does not have any information about the scene but causes haziness and low contrast.

1 Introduction

Blur, the reflected light from the object of interest to the camera is partially refracted and enters the camera at a slightly diverted angle which causes blurriness.

Noise, natural waters, both fresh and saline, are mixtures of dissolved and suspended matter. These solutes and particles are both optically significant and highly variable in kind and concentration. They not only reduce the scene visibility but also increase the aforementioned degradations.

These degradations are addressed via two image processing perspectives: image restoration and enhancement.

Image restoration techniques are the inverse problem which aim at recovering a distorted image. Doing this, usually, requires building mathematical models of the degradation. The quality of a restored image highly depends on the accuracy and validity of the degradation model. These methods can be very effective, though they require a significant number of parameters to characterize water turbidity and light attenuation coefficients.

In contrast, image enhancement methods seek to make the images more aesthetically pleasing through subjective criteria and without relying on complex mathematical models such as improving image contrast and brightness adjustment. They are not based on any physical model, therefore, they have less accuracy than image restoration approaches but are simpler and faster.

The research provided in this thesis focuses on restoring degraded UW images/frames using several proposed approaches. The methods are aligned with the line of image restoration dealing with the provided data. This means that, there is no control over the data acquisition. The proposed methods address UW image super-resolution, deblurring, and denoising by means of eliminating floating particles. The motivations are listed as follow:

1. Noting that transferring data is a very important issue especially in a UW situation, thus, the acquired images and videos are usually compressed. This means that they often suffer from low resolution.
2. Besides color cast and haze, strong blur due to light scattering in water, decreases image quality by smoothing the edges and fading the small details. This leads to poor performance of advance image processing algorithms such as segmentation and classification.
3. Floating particles in water are one of the main reasons to serve UW image degradations. These particles not only increase scattering and absorption effects but are also visually disturbing. Despite their important impact on UW image distortion, they can be very misleading in applications such as the surveillance of the subsea environment.

1.1 Thesis Contributions

The thesis contributions are detailed across different chapters and consequently, via different algorithms.

1. The first contribution is a new super-resolution and deblurring approach (Chapter 5). The main objective in this chapter is to devise a super-resolution approach via sparse coding over multiple class dictionaries. To do this, a specific classifier is needed to cluster the training data for dictionary learning. The training data in each cluster share a common feature differentiating them from other clusters. This results in much lower signal variability within a class compared to the signal space. This is one of the motivations for learning compact class dictionaries. Thereby, a pair of dictionaries is trained over each training set which results in improving the representation quality since each dictionary is concerned with a specific signal class. In addition, a deblurring algorithm is integrated into the super-resolution problem. This is the second motivation for using multiple dictionaries. Noting the fact that UW blur caused by scattering is dependent on camera-object distance, naturally, spatially variant, compact dictionaries corresponding to different spatial blurriness levels are required.

Thereby, a scale-invariant blurriness measure based on sparse representation over a degraded blur dictionary is proposed. This criterion is employed as the classifier in the dictionary learning phase and the model selection in the restoration stage. The main contributions made for the super-resolution and deblurring approach are:

- Introducing a blurriness measure (BM) based on the sparse representation of each patch on a blur dictionary. This measure is used to characterize spatial intensity variations of image patches and shown to be approximately invariant to the patch resolution. Therefore, it is used to separate image patches based on how blurred they are.
- Using the BM as a classifier, the training image patches are clustered and a pair of dictionaries is trained over each class. The dictionaries of each class suits its training cluster the best since the atoms of these dictionaries inherent a common feature (the same level of blurriness). A single image super-resolution approach based on sparse representation over the learned compact dictionaries is then introduced. In the restoration stage, the BM is applied as the model selection to restore image patches with the most appropriate dictionary pair.
- Furthermore, a deblurring approach is integrated into super-resolution method to handle the spatially variant UW blur caused by scattering. This results not only in restoration of HR images but also enhanced image details and structures according to their level of blurriness.

2. The second contribution Chapter UW image/video frame denoising via detection and removal of floating particles (Chapter 6). For this purpose, two algorithms dealing with

1 Introduction

single images and video frames are proposed. A novel filtering approach, termed Median², is suggested for single images. The algorithm detects each particle in the image and then, using the neighborhood pixels, estimates the scene behind them. The Median² is designed in a way that enables it to:

- differentiate between the marine particles and object edges.
- filter multiple corrupted pixels in a local patch at the same time without introducing the significant blurriness effect caused by traditional Median filter.
- reduce even the general blurriness effect due to filtering, meaning the image structures are preserved as much as possible.

As a result, it achieves the goals of filtering, removing the unwanted signal where the image structures and details are preserved. However, whenever there is too much noise, using only local information leads to some uncertainty or false restoration result, thus, a second algorithm, dealing with video frames, is proposed. The main assumption for this approach is that the videos are acquired with a static camera. However, the algorithm is adapted to the small movements of the camera such as pan, tilt and zooming, this includes the slow movement of the device as well (i.e. ROV).

The approach is conducted with four main steps, dividing the video frames into two categories (training and testing) first, a background model of the scene is obtained using the training frames. For this, a mixture Gaussian approach is introduced which adapts the background model to the current frame by constantly checking the foreground and background objects and any changes in their status. The approach is based on change detection, meaning the moving objects count as foreground and vice versa. Moreover, the obtained background model is modified to the camera movements by using the affine transformation of the movement. For this, a set of probable affine transformations are examined, with the one that matches the movement the best being chosen.

Second, now having the background model, some statistical priors of the scene are learned over this reference image. The background model provides the information about static components of the scene which contains the scene covered by marine particles. Third, employing the proposed Median² algorithm, a mask containing the locations of the particles is extracted from each test frame. To this end, instead of filtering the original image after detection of the particles, their location is marked in a mask image. Finally, each test frame is restored using its corresponding mask together with the statistical priors of the scene. This approach results in the following achievements:

- A higher restoration accuracy due to the fact that the prior knowledge learned from the whole scene, is used for restoration rather than using limited local neighborhoods.
- The blurriness effect further decreases since no filtering scheme is used.
- Any false detection and defect caused by filtering is fixed through employing the priors of the whole scene rather than just the neighborhood.

1.2 Outline of the Dissertation

This thesis contains six chapters, including this introduction, which is structured as follows:

Chapter 2 Underwater Imaging and Degradations

In this chapter, UW imaging and the degradations caused by the medium are first discussed. The well-known Jaffe-McGlamery image formation is then explained in detail.

Chapter 3 Literature Review

Here, a review of the existing works addressing the problems in UW images is provided. This is done in three main subsections, namely, image restoration, image enhancement and image denoising. At the end of this chapter, the literature is clustered based on approach methodology and application in different categories such as color correction, dehazing, deblurring and denoising. This helps the readers to find their approach of interest based on the application they are looking for.

Chapter 4 Sparse Representation: Definition and Applications

This chapter gives an insight into the definition of sparse coding and its formulation. A survey of existing approaches to atom selection problem is offered. The sparse representation problem needs the availability of a dictionary, thus, the possibilities for choice of dictionary are reviewed here and then approaches dealing with dictionary learning are supplied. Finally, the applications of sparse representation in the field of image processing are presented and the pioneering work relevant to this study, super-resolution, is discussed in detail.

Chapter 5 Underwater Single Image Super-Resolution and Deblurring via Sparse Representation

In this chapter a new method is proposed which integrates super-resolution and deblurring, especially for UW images. It is a dictionary learning-based approach which reverse the degradation model of single images to provide a high resolution and blur free output. The approach suggests use of multiple dictionaries, where a blurriness metric is proposed to cluster training data to train the proper dictionaries. The same metric is then used for the dictionary selection step at reconstruction stage. Super-resolution is applied only on HF components of input image where deblurring is applied according to the degradation level of image patches. It is demonstrated that the approach is capable of deblurring UW images when it increases the resolution of single input images, at least by a factor of 2.

This chapter is based on the following publications:

- Adaptive UW Image Deblurring via Sparse Representation, [FR16].

- Underwater Image Restoration: Effect of Different Dictionaries, [Far15].
- Learning-based underwater image enhancement with adaptive color mapping, [FZL15].

Chapter 6 Denoising: Marine Snow Detection and Removal

This chapter explores the definition of marine snow, its effect on imaging and the degradations it may cause. Next, this phenomenon is investigated in images to define its properties in the image domain. Two approaches addressing detection and removal of marine snow are then proposed. The first approach tackles single images where a filtering scheme, Median² is proposed for restoring the missing scene covered with marine particles. The second approach deals with videos where the camera is assumed to be static. In both methods, a detection algorithm is first offered to find the pixels that have been corrupted with marine snow. The results of both algorithms illustrate the success of the proposed methods at detecting and eliminating marine snow from single images as well as video frames.

This chapter is based on the following publications:

- Single Image Marine Snow Removal based on a Supervised Median Filtering Scheme, [FRL17b].
- Marine-Snow Detection and Removal: Underwater Image Restoration using Background Modeling, [FRL17a].
- Change Detection in Crowded Underwater Scenes - Via an Extended Gaussian Switch Model Combined with a Flux Tensor Pre-segmentation, [RFL17b].
- Background Modeling: Dealing with Pan, Tilt or Zoom in Videos, [RFL17a].

Chapter 7 Combined Super-Resolution, Deblurring and Marine Snow Removal

In this chapter, some examples of UW images are evaluated in a pipeline comprising both proposed algorithms.

Chapter 8 Conclusions and Future Work

A conclusion from the simulations and experimental results are drawn and possible future works are suggested.

2 Underwater Imaging and Degradations

2.1 Introduction

Physical properties of water cause degradation effects which are not present in images acquired in air. Light interacts with water and disturbs the captured image. Here a detailed explanation of those interactions and the resulting effects on an observed image are provided.

2.2 Underwater Effects

By cause of light interaction with water and its inherent particles, imaging systems give poor visibility results in this medium. This is due to the low energy received by the camera. Light loses its energy by going deeper in water or traveling through it. This happens due to scattering and absorbing properties of the transmitting medium.

Light travels from the source (sun) and interacts with water in three forms, refraction, scattering and absorption:

- Refraction occurs when light hits an interface between air and water non-perpendicularly. The light ray r in water has a different direction than the incident ray i . The angle of refraction β is smaller than the angle of incident α since the propagation speed of light in water is lower than in air, so, the refracted ray r is refracted towards the interface normal s (Figure 2.1(a)). As a result, submerged objects seem to be closer to the observer and hence bigger than they are in reality. It is dependent on the distance of the imaged objects to the refractive interface and the incidence angle of light rays entering the camera through this interface.
- Scattering changes the direction of light from its way. This happens in two scenarios:
once the light enters water, a fraction of it reaches the object. This light is then reflected to the camera, however, due to the scattering properties of water, a portion of it is scattered with small angles. In underwater imaging terminology this is called "forward scattering" which is light deviation from the object to the camera, with this generally leading to blurring of the image features (red rays in Figure 2.1(a)).

On the other hand, a part of the light source never hits the object but instead it is reflected back to the camera by water molecules or floating particles that are in the water (blue ray in Figure 2.1(a)). This, what is know as, "backscattering" yields a degraded contrast scene

2 Underwater Imaging and Degradations

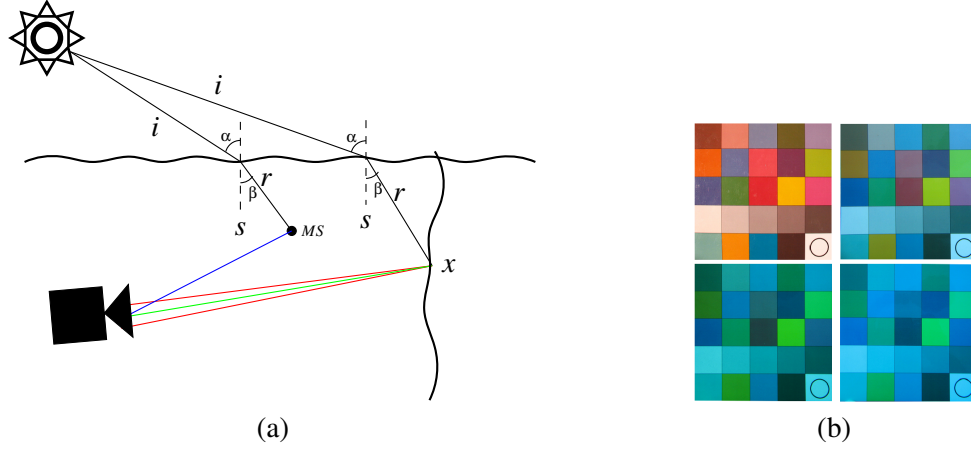


Figure 2.1: Illustration of light interaction with water. (a) Three components of the light paths affecting the perceived pixel intensity. The direct component (green ray), forward scattered component (red rays) and backscattering component (blue ray). (b) Light absorption, top-left: water surface, top-right: 5m depth, bottom-left: 10m depth, bottom-right: 20m depth.

and a foggy appearance by generating a characteristic veil that superimposes itself on the image and hides the scene. Distant objects therefore, appear misty.

Based on the Rayleigh scattering theory [Hah06], the intensity of scattering is inversely proportional to the fourth power of its wavelength, so that colors with shorter wavelengths such as violet and blue light are scattered much more than the longer wavelengths of yellow and especially red light.

- Absorption reduces the light energy. The amount of sunlight decreases depending on depth, while colors disappear one by one based on their wavelengths. This leads to one color dominating an image. According to the selective absorption of water, colors with a longer wavelength are much easier to be absorbed so red light will be absorbed before colors with shorter wavelengths such as blue and green (Figure 2.1(b)).

In this work, both the camera and the scene are in the same medium (underwater), whereas, the light journey from the scene to the camera is of particular importance, thus, scattering and absorption are the main focus rather than refraction.

In conclusion, water absorbs the longer wavelength such as the red color and scatters the shorter wavelength such as blue and violet colors when visible light disseminates in it. Practically, at a certain distance, about twenty meters in clear water and five meters or less in turbid water [Baz+06], objects are almost indistinguishable while the colors are faded. However, the visibility range can be increased using artificial illumination of the object, but it produces a non-uniform light on the surface of the object as well as a bright spot in the center of the image with a poorly

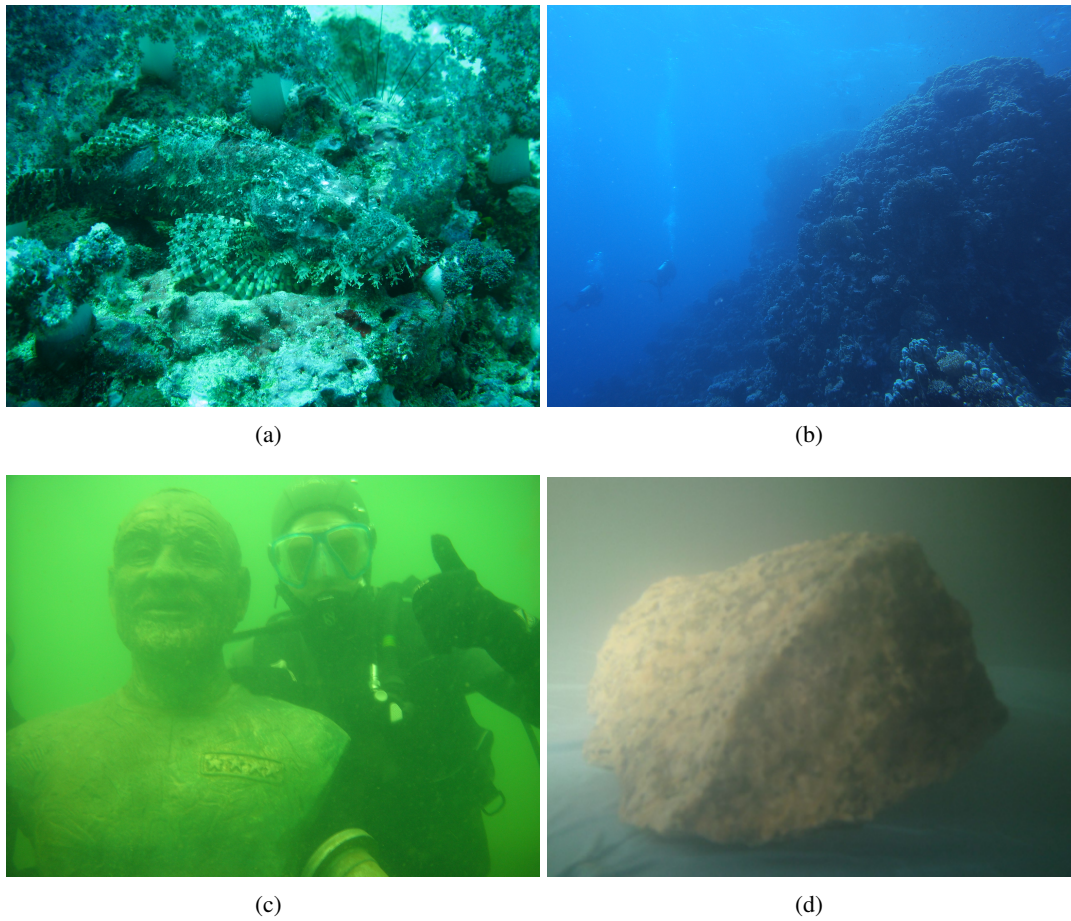


Figure 2.2: Four examples of underwater images suffering from (a) diminished colors, (b) low contrast and haze, (c) blur and color cast and (d) blur. Images are courtesy of Eik Deistung (a)(b), Ancuti et al. [Anc+12] (c) and Wu et al. [WL13] (d).

illuminated area surrounding it.

Hence, underwater images usually suffer from at least one of the following challenges: low contrast, significant blur, limited range (haze) and diminished color. Figure 2.2 shows four underwater images taken in different locations which suffer from different degradations.

2.3 Underwater Image Formation

Imaging is limited by the environment and the technologies that are currently employed. These limitations can be determined by two factors: energy and contrast. The energy which carries the information about the object is reduced by the light absorption. It is a function of depth and medium refractive index. On the other hand, contrast is limited by the light scattering. It is related

2 Underwater Imaging and Degradations

to the medium refractive index, its volume and illumination source.

In order to model a captured image, it is crucial to have an estimation of how light propagates while traveling. Lambert-Beer took into account the above-mentioned factors and empirically defined an equation to represent the decay of light intensity as follows:

$$E(r) = E(0)e^{-cr}, \quad c = a + b \quad (2.1)$$

here r is distance, E is the irradiance at the positions 0 and r , and lastly, c is the total attenuation coefficient of the medium. The constant c is defined as the sum of two quantities, the absorption coefficient a and scattering coefficient b . Using this fundamental knowledge about the light interaction with the medium, computational modeling approaches are proposed to model the captured image from the perspective of a camera underwater.

Duntley [Dun63], accomplished pioneering work in underwater optics including the optical nature of light, the penetration of daylight into the sea. The results of this research, which was initiated in the 1950s and 1960s, was pursued in the 1970s by McGlamery [McG80]. McGlamery laid out the theoretical foundation of the optical image formation model which was extended in the 1990s by Jaffe [Jaf90] to design different subsea image acquisition systems. This image formation is the basis of many underwater image restoration algorithms, thus, it is explained in more detail in the following section.

2.4 Jaffe-McGlamery Model

McGlamery [McG80] developed a model to formulate the outcome of underwater imaging systems. He claimed that the image plane irradiance consists of three components: direct component, forward scatter and backscatter. They can be clustered into two main groups: irradiance due to the light reflection from the object, non-scattered light (direct component) and scattered light (forward scatter), and the irradiance which is the reflected light by the water molecules and floating particles (backscatter).

Thus, the total irradiance received by a pixel p at the position (x, y) in the image plane, termed $E_T(x, y)$, is the linear superposition of three aforementioned components (Figure 2.1(a)).

- The direct component $E_d(x, y)$ is a high resolution image formed by the reflected light from the object to the camera without being scattered (green ray).
- The forward scatter component $E_{fs}(x, y)$ is a degraded image formed by reflected light from the object to the camera via small angle scattering processes (red rays).
- The backscatter component $E_{bs}(x, y)$ is the reflected light from the source which is scattered towards the camera by the water volume in front of the camera (blue ray).

Therefore, an underwater image is defined as follows:

$$E_T(x, y) = E_d(x, y) + E_{fs}(x, y) + E_{bs}(x, y) \quad (2.2)$$

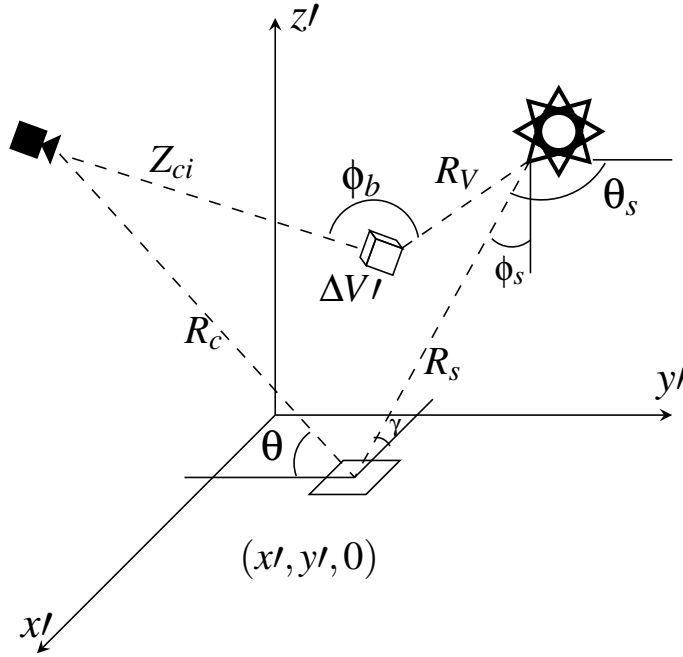


Figure 2.3: Coordinate system of Jaffe-McGlamery's model.

2.4.1 The Direct and Forward-Scattered Components

As mentioned, the direct and forward scatter components are due to the reflected light from the object. Thus, to calculate these components, the irradiance incident upon the reflectance map which originates from the light source must be calculated. Accordingly, the light journey from the source to the object and further from the object to the camera aperture must be explored. Based on the proposed method by Jaffe [Jaf90], which is closely related to the McGlamery's approach, this process can be divided into three sub-procedures:

- light from the source to the object, which is attenuated and spherically spread before reaching the object. The light source assumed to be a point-source with irradiance $BP(\theta_s, \phi_s)$ on a hemispherical shell at a unit distance (1 m) from the light source, a function of polar angles θ_s and ϕ_s (see Figure 2.3). Thus, the incident irradiance, on the object, with a fixed coordinate system $(x', y', z' = 0)$ consists of the spreading of illumination due to a small-angle forward scattered component as well as adding back in a contribution due to small-angle forward scattering as follows:

$$E_I(x', y', 0) = E'_I(x', y', 0) * g(x', y' | R_s, G, c, B) + E'_I(x', y', 0), \quad (2.3)$$

where

$$E'_I(x', y', \theta_s, \phi_s) = BP(\theta_s, \phi_s) \cos \gamma \frac{e^{-cR_s}}{R_s^2}, \quad (2.4)$$

here R_s is the distance from the source to the object, c is the attenuation coefficient and γ is the angle between a perpendicular to the given location x', y' and the line between

2 Underwater Imaging and Degradations

source to the x', y' location. Note that " $*$ " operator denotes convolution. The Point Spread Function (PSF), $g(x', y' | R_s, G, c, B)$, is defined via an experimental tank filled with sea water [McG80] as:

$$g(x', y' | R_s, G, c, B) = (e^{-GR_s} - e^{-cR_s}) \mathcal{F}^{-1}(e^{-BR_s f}), \quad (2.5)$$

here G is an empirical constant ($|G| \leq c$), B an empirical damping factor and f is spatial frequency in *cycles/rd*. The operator \mathcal{F}^{-1} represents an inverse Fourier transform.

- The reflected light from the object to the camera. The reflected radiance is then computed by multiplying the incident irradiance pattern by the reflectance values of the reflectance map. The reflectance values of the reflectance map for oceanographic objects are $0.02 \leq M(x', y') \leq 0.1$ ([Dix+83]), where $M(x', y')$ is defined as a planar reflectance map located in a three dimensional space at a fixed orientation to represent an object.
- The attenuated light based on Lambert law, irradiance on the image plane is less than the radiance that enters into the camera. The attenuation model is defined as:

$$\frac{\cos^4 \theta T_l}{4f_n} \left[\frac{R_c - F_l}{R_c} \right], \quad (2.6)$$

the angle θ is the one between the reflectance map and the line between the position x', y' and the camera aperture (Figure 2.3). Here R_c is the distance from a position x', y' on the reflectance map to the camera, f_n is the f number of the camera of focal length F_l , and T_l is the transmittance of the lens.

Therefore, the direct irradiance (E_d) incident upon the image plane of the camera is represented as follows:

$$E_d(x, y) = E_l(x', y', 0)(e^{-cR_c})M(x', y') \cdot \frac{\cos^4 \theta T_l}{4f_n} \left[\frac{R_c - F_l}{R_c} \right]^2. \quad (2.7)$$

Note that the primed coordinate system stands for the reflectance map while the unprimed coordinate system is associated with the camera plane.

Once the direct component is calculated, the forward-scatter (the second component) can be calculated from the direct component via convolution with the Point Spread Function (PSF) $g(x, y | R_c, G, c, B)$:

$$E_{fs}(x, y) = E_d(x, y, 0) * g(x, y | R_c, G, c, B), \quad (2.8)$$

where g is represented by 2.5.

2.4.2 The Backscatter Component

In order to calculate the third component (backscatter), the linear approximation used for forward-scattered component computation, small angle theory [Wei73], is no longer valid since backscattered light enters the camera from a large distribution of angles.

For this, they consider that the three dimensional space is sliced into planes of thickness ΔZ_i that are parallel to the image plane of the camera (Figure (2.3)). Each slice is then treated the same as the above argument and its irradiance incident is calculated by computing the direct component and adding the additional irradiance due to the forward scattered component. Thus, the irradiance incident on the volume elements is as follows:

$$E_s(x', y', z') = E_{s,d}(x', y', z') + E_{s,fs}(x', y', z') \quad (2.9)$$

where $E_s(x', y', z')$ is the irradiance in three dimensional space propagating away from the light source. Next, the irradiance incident upon the camera element (x, y) due to volume $\Delta V'$ is defined as:

$$E_{bs}(x, y) = E_{bs,d}(x, y) + E_{bs,d}(x, y) * g(x, y | R_c, G, c, B). \quad (2.10)$$

where $E_{bs,d}(x, y)$ is the direct component of the backscattered irradiance and is defined as follows:

$$E_{bs,d}(x, y) = \sum_{i=1}^N e^{(-cZ_{ci})} \beta(\phi_{bs}) E_s(x', y', z') \cdot \frac{\pi \Delta Z_i}{4 f_n^2} \cos^3 \theta T_l \left[\frac{Z_{ci} - f_l}{Z_{ci}} \right]^2 \quad (2.11)$$

where ΔZ_i is the thickness of backscattering volume $\Delta V'$, Z_{ci} is the distance from a point in the camera to the center of the backscatter slab and N is the number of the backscatter plane.

Once all three components are calculated, the underwater image is modeled using Equation 2.2. This comprehensive model is often used for image restoration approaches. In the next section, a short summary of some of those methods is given.

3 Literature Review

3.1 Introduction

Capturing a clear scene UW is not a trivial task but important, since the quality of these images play a crucial role in success/failure of applications such as UW navigational monitoring and environment evaluation. This issue can be addressed from two perspectives: hardware and software. Using advanced technologies such as a high-end camera can reduce the degradation effects, however, the images still suffer from low visibility and are not comparable to in air images. Therefore, UW image processing algorithms have attracted a lot of attention in terms of providing sufficient quality of images where there is no need for expensive hardware. For this, two main approaches are introduced: image restoration and enhancement.

Image restoration techniques are the inverse problem which aim at recovering a distorted image using the degradation model. The quality of restored image highly depends on the accuracy and validity of the degradation model. These methods can be very effective but they require a significant number of parameters to characterize water turbidity and light attenuation coefficients.

In contrast, image enhancement methods improve the visibility of one or more aspects of an image such as sharpening of the image features or improving image contrast. They are not based on any physical model, therefore, they have less accuracy than image restoration approaches but they are simpler and faster.

In this report, the most pioneering, state-of-the-art and recent underwater image restoration and enhancement techniques are reviewed. Furthermore, a section is considered for UW image denoising. Image denoising is a common restoration approach however, due to the special case of floating particles UW, this issue is addressed in a separate section. Floating particles made of mostly organic matter are considered as unwanted signal since they decrease image visibility where also disturb image processing and computer vision tasks. This signal is considered as noise and is addressed in UW image denoising algorithms together with other common imaging noises.

At the end of this chapter, the reviewed approaches are classified based on the main objective (deblurring, color and contrast enhancement, dehazing, and denoising), the methodology and the evaluations provided in each piece of literature.

3.2 Underwater Image Restoration

A general model to represent an observed image $I(x, y)$ is defined as:

$$I(x, y) = J(x, y) * h(x, y) + n(x, y). \quad (3.1)$$

Image restoration aims at recovering the clear image $J(x, y)$ from the observed image $I(x, y)$ taking into account the explicit knowledge of the point spread function $h(x, y)$ in the presence of noise $n(x, y)$. Therefore, having an estimation of PSF together with the noise, the clear image can be restored. The PSF of degradation components includes imaging system response and effects of a medium. The better the available knowledge about the degradation function is, the better results can be obtained from the restoration.

The next two subsections summarize the existing methods dealing with UW image restoration by:

- estimating the PSF of the degradation components of UW image formation.
- using the statistical prior.

3.2.1 Image Restoration based on Point Spread Function Estimation

Here the methods which apply the geometric properties of light propagation and propose restoration algorithms are introduced. The Jaffe-McGlamery model is the basis for many of these algorithms. Although, due to the simulation complexity of this model, later approaches look at its components separately. They provide the corresponding solution by simplifying the model.

Schechner and Karpel [SK05] proposed an image restoration based on polarization effects in UW scattering to compensate for the poor visibility in this medium. They exploited the effect of each degradation source using Jaffe-McGlamery image formation simulation and claimed that the main cause of contrast deterioration is veiling light (backscatter) rather than the blur (forward-scattered).

Therefore, the low contrast and color diminished are considered as the main concern and the blur effect is discarded. They developed the method based on the fact that backscattered light under natural UW lightening is partially polarized horizontally. For this, they mounted a Polaris on a camera and took several raw images at different orientations to sense different polarization components of the veiling light. Even when the raw images have very low contrast, their slight difference between the extremum transmittances is used as the key for visibility improvement.

In addition, an algorithm is used to enable proper compensation for the color bias. They used a contrast criterion as a logarithmic function of the backscatter component to evaluate their algorithm quantitatively. Furthermore, the approach automatically accounts for dependencies on object distance, so as a by-product a distance map of the scene is obtained. Experiments illustrate that the method is able to restore the contrast and colors of UW images, nearly doubling the visibility range.

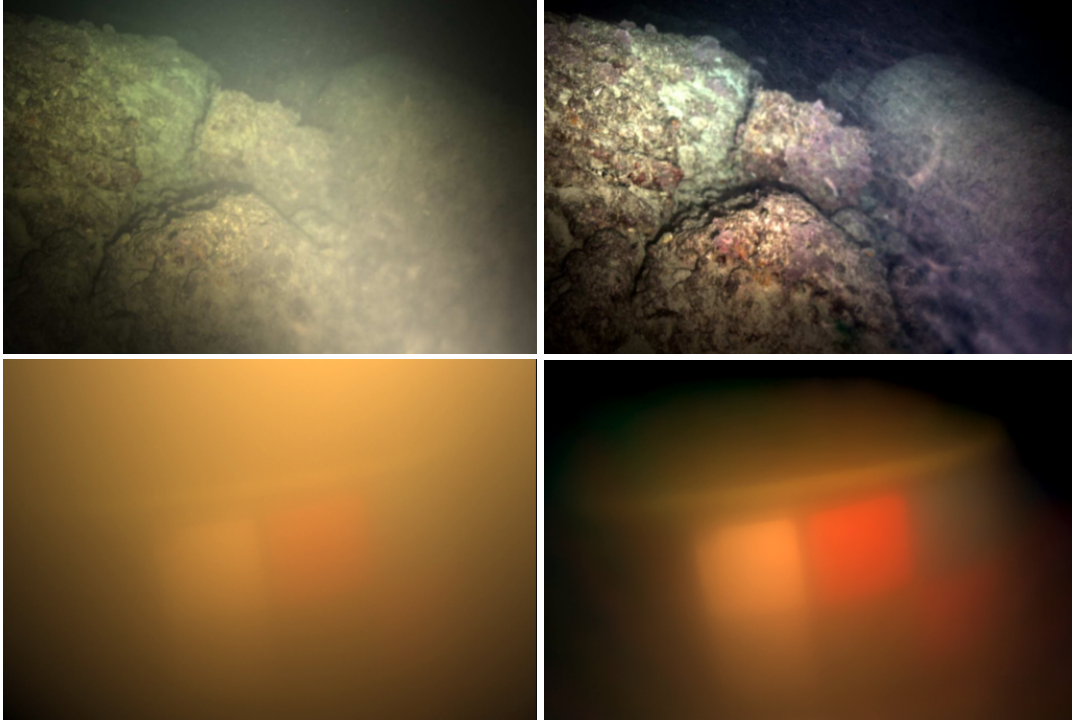


Figure 3.1: UW scenes at the Mediterranean sea and the sea of Galilee from top to bottom respectively. From left to right: the raw image and the result after backscatter removal. Image courtesy of Treibitz and Schechner [TS09].

The same logic was followed later by Treibitz and Schechner [TS09]. To provide a more general solution, they took into consideration not only the backscattered light but also the object reflection to be partially polarized. They studied the image formation in the presence of artificial illumination. The Polaris was mounted on the artificial light and an analyzer on the camera and only two frames of the scene were taken with different states of the analyzer or Polaris. The method is simple and gives the visibility recovery while providing a 3D estimation of the scene structure. The results show the success of de-scattering of images taken in various sea environments at night. Some of the results for this technique are shown in Figure 3.1.

In the above-mentioned works, blurring was considered as the least disturbing, thus it was ignored in the restoration process. To address this difficulty, Trucco and Olmos-Antillon [TO06] proposed a self-tuning image restoration filter which aims at recovering blurred images. They simplified UW image formation by acknowledging the forward scattered as the major degradation and discard the explicit expression of the backscattered component. It seems to be reasonable whenever the concentration of particulate matter generating backscatter in the water is limited. For this, they assumed that illumination is uniform all over the image, such an assumption is acceptable for shallow water where illumination is provided by direct sunlight. The filter is designed by taking into account only the forward scatter component as the main degradation



Figure 3.2: UW scene from Elba, Italy. Left is the input image and right is the restored image. Image courtesy of Trucco and Olmos-Antillon [TO06].

factor. As a result, an inverse filter is designed in the frequency domain as follows:

$$S(\omega) = \frac{1}{K} e^{cR_c \omega}, \quad (3.2)$$

where K is an experimental constant for the difference of exponentials in Equation (2.5). In order to estimate the optimal values for the filter, they applied a quality criterion based on a global contrast measure and minimized it over each individual image. The optimal environment for this method is shallow water with low backscatter light. Figure 3.2 shows an example of the results using this filter.

This simple still effective approach was followed by Wu and Li [WL13] to give a more accurate solution towards deblurring. Trucco and Olmos-Antillon [TO06] only considered the forward scatter component and ignored the direct component and their relation which is the reason for why the image is low contrasted. This was the motivation for Wu and Li to provide a refined inverse filter. The same as [TO06], they focused on how scattering affects the imaging system rather than absorption. Since backscattered light does not carry information about the scene, they did not consider this as part of the model and addressed the problem as:

$$E_T(x, y) = E_d(x, y) + E_{fs}(x, y). \quad (3.3)$$

To optimize the model, they assumed that water consists of many slices parallel to the image plane with ΔZ thickness. Further simplification was done by considering that each water layer scatters the same amount of light as the layer behind, meaning no light is absorbed by the layers. Thus, they integrated the results over all the layers and designed an inverse filter in the frequency domain as:

$$S(\omega) = \frac{1}{1 + k \frac{1 - e^{-b\omega}}{\omega}}, \quad (3.4)$$

where $b = BR_c$ and k is derived from PSF function in Equation (2.5). In order to estimate the parameters in the model, they used an image quality measure and optimized it over the entropy of the restored image in an iterative manner. The method gives good quality in the case of deblurring while some of the lost information can be recovered successfully (Figure 3.3).

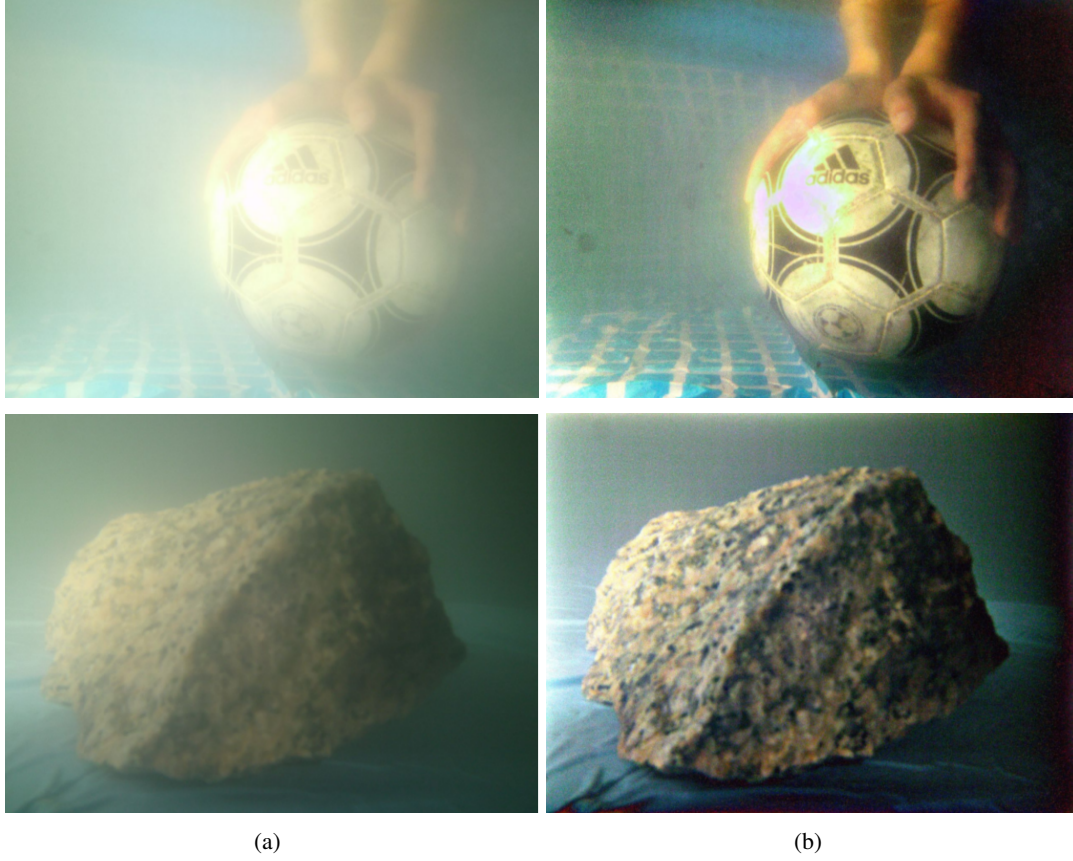


Figure 3.3: UW scenes in an artificial pool. (a) Input image and (b) restored image. Image courtesy of Wu and Li [WL13].

Another possible approach for UW image restoration is to apply the traditional image restoration framework while the PSF of the water body is calculated by simulation or modeling. Liu et al. [Liu+01; ZTG03] proposed a restoration of blurry UW images based on Wiener filtering. For this they used a remotely operated vehicle (ROV) to capture videos of a standard target through a 1.5 m water body. The captured data is then used to calculate the PSF. They simulated the maximum detection range of the ROV under different water body conditions and calculated the captured irradiance. Once the PSF of the water is available, they recovered blurry images using the results of simulation together with a Wiener filter.

Inspired by [Liu+01; ZTG03], Hou et al. [Hou+07a; Hou+07b; Hou+08] reported an automated image restoration termed NRL Image Restoration via Denoised Deconvolution (NIRDD) which tackles blurring through strong scattering. The UW optical properties are incorporated into the PSF in the spatial domain and the modulation transfer function (MTF) in the frequency domain. The total system response, the MTF term $H(u, v)$, is the effect of multiple individual components,

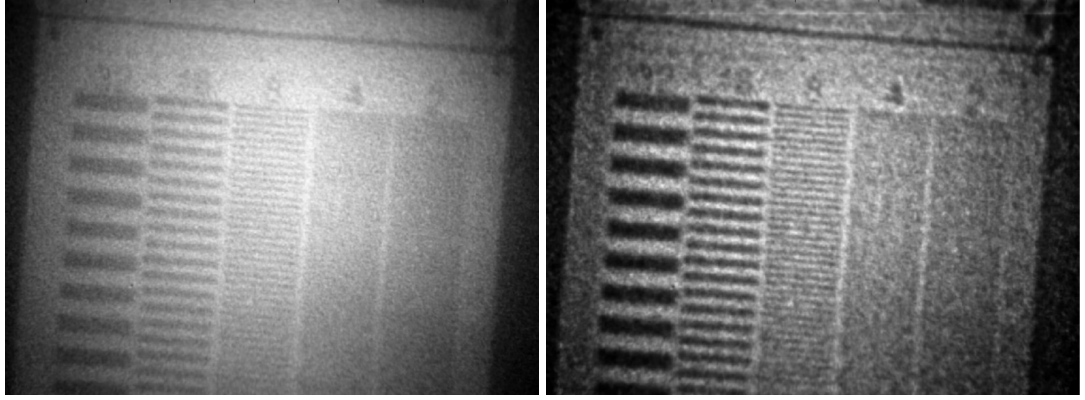


Figure 3.4: UW scene taken at 7.5 m depth in Florida. From left to right: original image and restored image based on measured PSF. With WGSA values of 0.05 and 0.14 respectively. Image courtesy of Hou et al. [Hou+07b].

i.e. the medium and the optical system, so can be expressed as follows:

$$H(u, v) = H_{\text{system}}(u, v)H_{\text{medium}}(u, v) \quad (3.5)$$

According to the small angle approximation [Wel73], the MTF of the medium is expressed as:

$$H_{\text{medium}}(\Psi, r) = e^{-D(\Psi)r} \quad (3.6)$$

where $D(\Psi)$ is the decay transfer function (DTF) and is independent of the range of detection and defined as:

$$D(\Psi) = c - \frac{b(1 - e^{-2\pi\theta_0\Psi})}{2\pi\theta_0\Psi} \quad (3.7)$$

here θ_0 is related to the mean square angle (MSA), b is the total scattering coefficient, and c is total attenuation coefficient.

Moreover, they suggest that the PSF of the medium can also be derived via Monte Carlo simulation. Finally, the restoration is carried out using the modeled or the measured medium PSF together with the system PSF carried out with knowledge of camera/lens MTF.

Besides qualitative evaluation, an objective quality metric, weighed grayscale angles (WGSA) [HW07], is used to evaluate the results. This metric is used to automate the approach where the total sharpness of the image before and after restoration is compared in order to achieve the best restoration quality. WGSA gives a value for the total sharpness of grayscale images by looking at their edge profiles. A pair of images before and after applying this approach is shown in Figure (3.4).

To cover both forward and backscatter effects, Wang et al. [WZF11] proposed an approach which allows for the PSF caused by both forward scatter and backscatter components. To estimate the PSF, they considered the scattering volume as many thin layers. The light transmission in the water is divided by n layers with sufficiently small thickness. Each layer is considered to

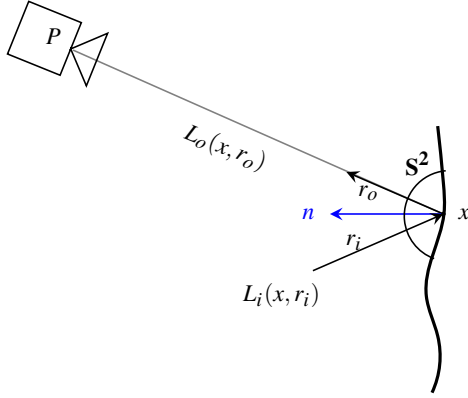


Figure 3.5: Radiance $L_o(x, r_o)$ reflected by the object at point x towards the camera can be calculated via incoming radiation $L_i(x, r_i)$ 3.2.1.

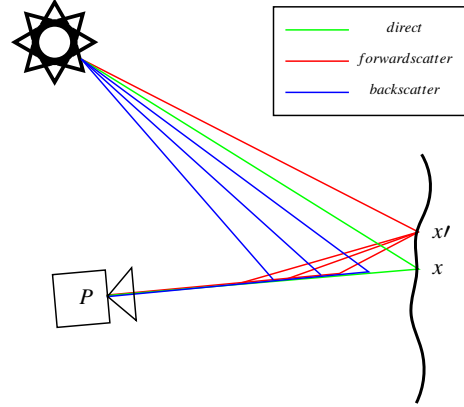


Figure 3.6: Three components of the light paths affecting the perceived pixel intensity proposed by Roth and Beyerer [SB14].

generate forward scattering and backscattering whenever light passes through it. Applying *in situ* measurement of the backscattering background, they estimated the parameters for the restoration. Finally, they applied the Wiener filter to achieve the restored image. For a single layer with sufficiently small thickness, the scattering particles are assumed to be statistically independent so the noise for backscattering is assumed to have a zero mean. Furthermore, the sensor transfer function is ignored in this model, thus assuming it to be ideal.

The Jaffe-McGlamery [Jaf90; McG80] model assumes the phase function to be small-angled to calculate the PSF function, thus, applying this assumption, they heuristically modeled the backscatter component. This implicit formed model cannot be easily used for image restoration. Counting this as a drawback, Stephan and Beyerer [SB14] proposed a computer graphical model based on a recursive rendering equation [ICG86; Kaj86] to fulfill the specific properties of UW imagery components, $E_d(x, y)$, $E_{fs}(x, y)$ and $E_{bs}(x, y)$.

According to the recursive integral equation, the radiance L_o leaving a point x into direction r_o is given by the sum of emitted radiance L_e and ingoing radiation L_i (Figure 3.5):

$$L_o(x, r_o) = L_e(x, r_o) + \int_{S^2} \rho(x, r_i, r_o) L_i(x, r_i) \langle r_i, n \rangle dr_i \quad (3.8)$$

where $r \in S^2 = \{r \in \mathbf{R}^3 \mid \|r\| = 1\}$ is a normalized direction, $\rho(\cdot)$ is the bidirectional reflectance distribution function depending on the direction of ingoing radiance r_i and outgoing radiance r_o and $\langle r_i, n \rangle$ is the cosine angle of the ingoing radiance and the normalised surface normal n .

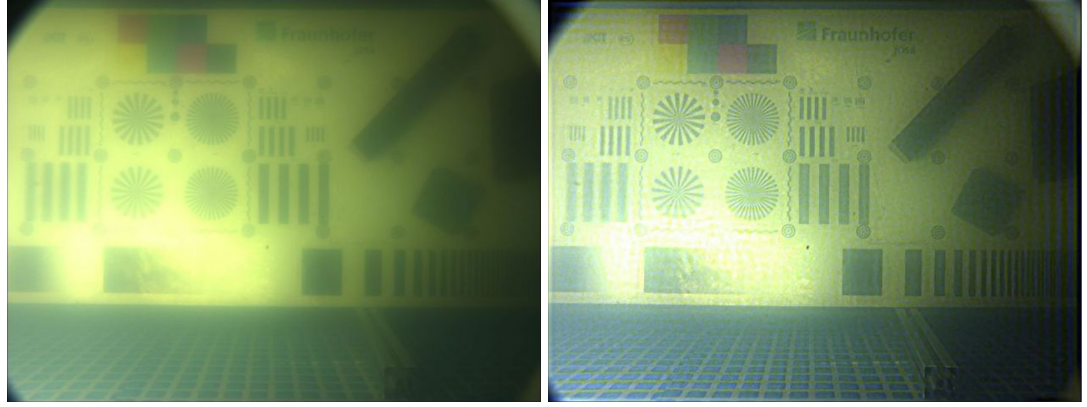


Figure 3.7: UW scene taken from a target image in an artificial pool. Original image (left) and the restored image using the proposed approach (right). Image courtesy of Stephan and Beyerer [SB14].

Employing this definition, they defined each UW imaging component and proposed a single scattering model to explain the low intensities, loss of contrast, color shift, brightening by backscatter and blur. For this, the camera is modeled as a pinhole and the medium is assumed to be homogeneous. Figure 3.6 illustrates the three components of the model. The direct component (green ray) represents the path of light reflected by the object surface point x towards the camera pinhole P . The forward scatter component (red rays) contains all paths reflected by an object surface point x' which are scattered into the line of sight \overline{Px} , which leads to the blur of the feature point x . The backscatter component is defined by the blue rays which represent the light rays emitted by the light source and scattered into the line of sight \overline{Px} without being reflected on the scene surface. It introduces low contrast and brightens the scene.

Once all three components are calculated via the recursive rendering, the image I is represented by means of affine transformation of reflectance vector \mathbf{p} and transfer function:

$$I = \Gamma \mathbf{p} + b, \quad \text{where} \quad \Gamma = \Gamma_d + \Gamma_{fs} \quad (3.9)$$

here b is the backscatter component which does not interact with the object surface reflectance. Furthermore, they explored the reliability of the model under the restriction of single scattering assumption. For this, the *optical depth* is used which is the stochastic mean path length that a photon travels without being scattered or absorbed [Mob94]. They illustrated that their model is valid for UW vision tasks at a minimum up to one optical depth where multiple scattering has only a slight impact on the imaging processes. Figure 3.7 demonstrates one example of the result where the proposed method is applied.

3.2.2 Image Restoration Using Statistical Prior

Many developed color and image restoration techniques are based on a model widely used to describe the formation of a haze image. This model was originally proposed to remove haze from

outdoor images, it considers a color observed image $I(x, y)$ composed of two components: the direct transmission of light from the object to the camera and the transmission due to scattering by the medium's particles [DBP57] as follows:

$$I(x, y) = J(x, y)t(x, y) + A(1 - t(x, y)), \quad (3.10)$$

where $\lambda \in (\text{red}, \text{green}, \text{blue})$, $J(x, y)$ is the scene radiance, $A(1 - t(x, y))$ is the *airlight* [Kos24] and refers to the ambient light reflected into the line of sight by atmospheric particles. The transmission $t(x, y)$ is defined as:

$$t(x, y) = e^{-bd(x, y)}, \quad (3.11)$$

here $d(x, y)$ is the scene depth and b is the scattering coefficient and is constant for a given pixel over all three channels. $I(x, y)$, $J(x, y)$ and the global atmospheric light A are all vectors in \mathbf{R}^3 . Assuming to have an estimation of A , to recover the clear image $J(x, y)$, the transmission $t(x, y)$ which captures the depth of scene is needed. However, the problem is under-constrained if the input is only a single haze image. Without additional knowledge and an assumption about the scene, it is not a trivial task to deduce if the color of a patch is due to the scene depth or by the natural color of the object. To overcome this ambiguity, often a statistical prior such as a "dark channel" [HST09; HST11] is used.

Dark channel is based on the observation that, in a haze-free outdoor image where the image patch is not "background light" (sky), some pixels have a very low intensity at least in one color (rgb) channel. It means that, the minimum intensity in such a patch should be of a very low value, i.e., a dark channel. The intensity of these pixels in a haze image is mainly contributed by the airlight, therefore, they can provide an accurate estimation of the haze's transmission.

Using the above-mentioned approach and adapting it for a UW situation, plenty of papers have addressed UW image dehazing and color restoration.

Carlevaris-Bianco et al. [CME10] employed the above mentioned image formation together with the UW model proposed by Duntley [Dun63] and introduced a dehazing algorithm specifically designed for UW images. In order to estimate the depth map, authors offered a simple and effective prior which exploits the wavelength-dependent attenuation of light for three color image channels. For this, they calculated the difference between the maximum of red channel intensity and the maximum of the green and blue channel intensities over a small image patch Ω as a function $D(x, y)$. Then the transmission function $t(x, y)$ in Equation 3.11 is estimated by shifting the values of D so that the largest difference between color channels, which represents the closest point in the foreground, is one.

The coarse initial estimation is further refined by applying the natural image matting algorithm [LLW08]. The so-called background light (airlight in the original definition) is then calculated using transmission estimation, the pixel with minimum transmission estimation represents the furthest point from the camera which is considered as the background light estimation. Once the depth map is estimated, the clear image is determined directly from the observed image using the

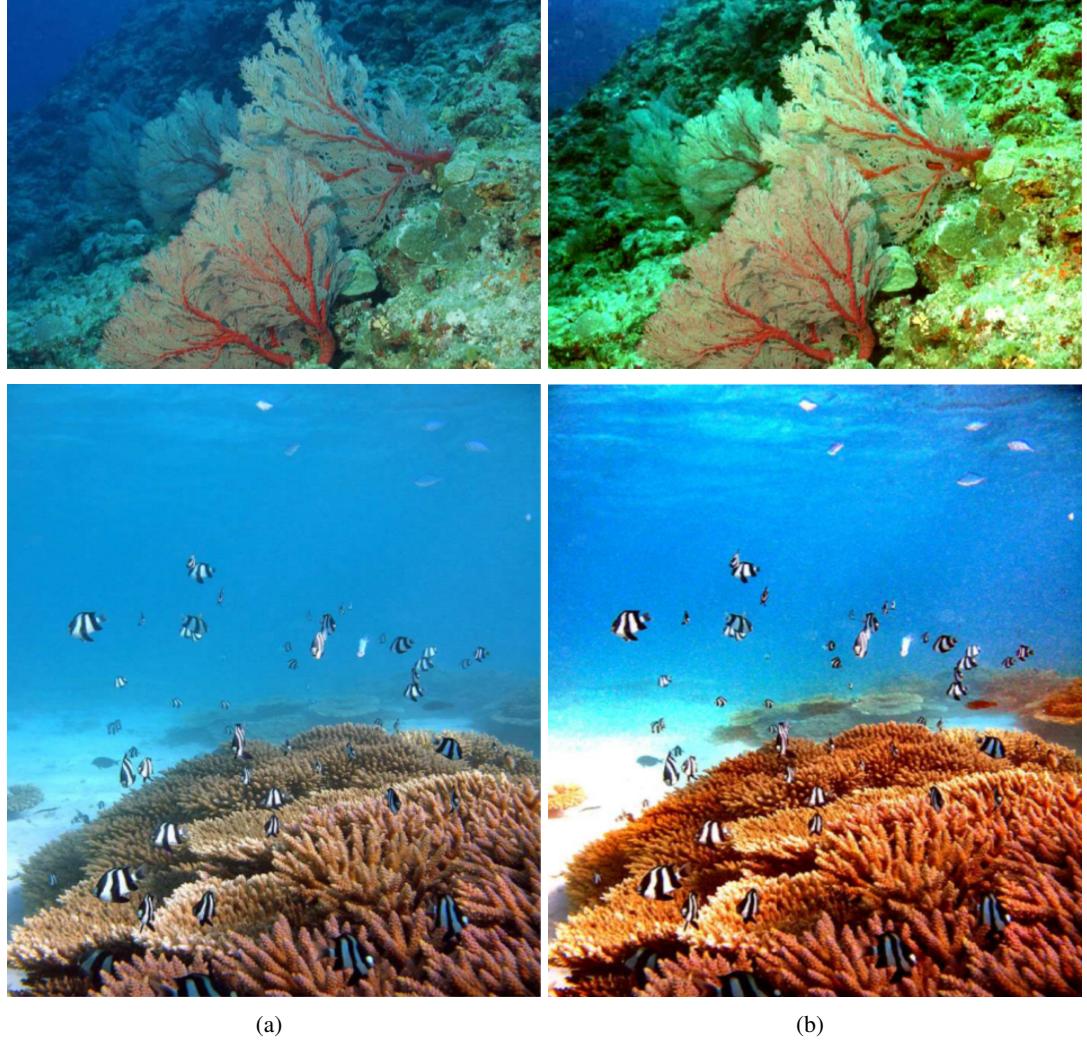


Figure 3.8: UW scenes of natural environment. Original images (a) and enhanced images (b). Image courtesy of Carlevaris-Bianco et al. [CME10].

following equation:

$$J(x,y) = \frac{I(x,y) - A}{t(x,y)} + A. \quad (3.12)$$

Moreover, they regularized the scene radiance calculation using a model that consists of Markov Random Field (MRF) under white Gaussian noise. To visually evaluate the approach, some of the results are shown in Figure 3.8.

As it is known, haze phenomenon is due to backscattering and is a function of distance. But not only object-camera distance but also depth range of photography is an important factor to increase the haze which was not addressed in the previous work.

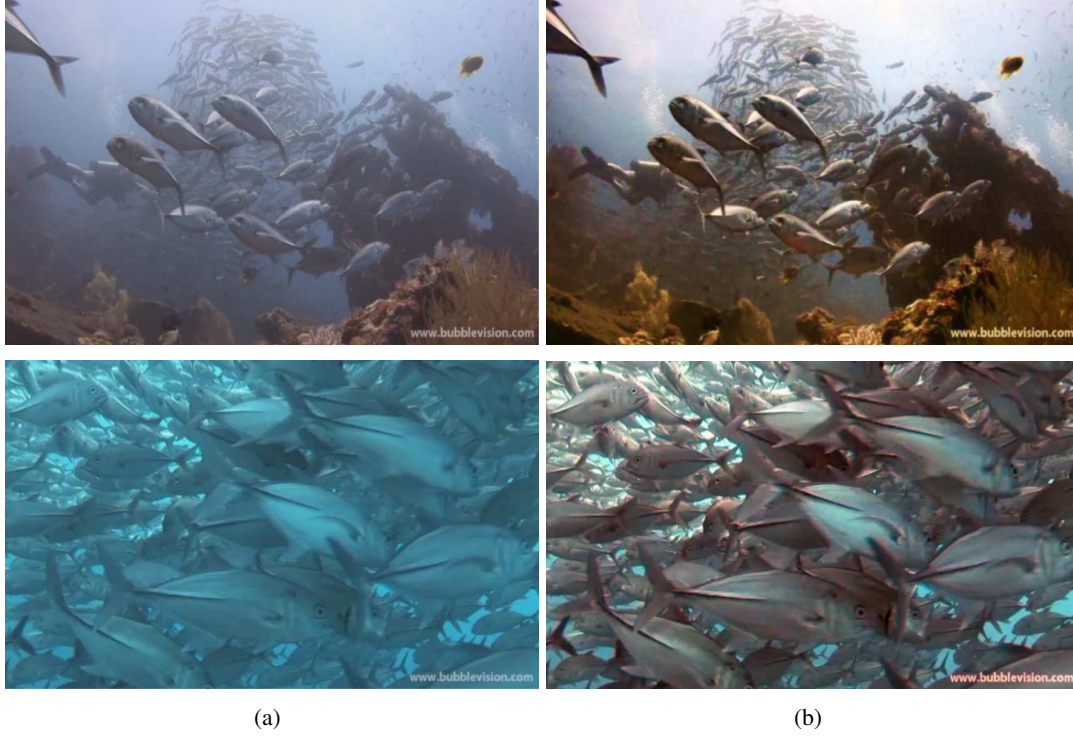


Figure 3.9: UW scenes taken in depth range of 18-22 m in natural environment. (a) Original image (b) dehazed result after applying WCID. Image courtesy of Chiang and Chen [CC12].

Chiang and Chen [CC12] presented a comprehensive method concerning dehazing UW images by considering three different depth maps: object-camera distance, depth at the top of photographic scene and image depth range. The approach is called wavelength compensation and image dehazing (WCID) which combines dehazing with wavelength compensation in order to address the distortion from color scatter and color cast [CC12]. First, they applied dark channel prior [HST09; HST11] to estimating the object-camera distance:

$$J_{\text{dark}}(x_c, y_c) = \min_{\lambda \in \{\text{red}, \text{green}, \text{blue}\}} \left(\min_{(x, y) \in \Omega(x_c, y_c)} J_{\lambda}(x, y) \right), \quad (3.13)$$

where $J_{\lambda}(x, y)$ is the observed UW haze image and Ω is the image local patch centered at (x_c, y_c) . Once the distance between object and the camera $d(x, y)$ is derived, they removed the haze and compensate the color cast encounter during the object-camera path. Furthermore, the photographic depth at the top of the scene and depth range of the scene are approximated from residual energy ratios of each wavelength in the background light of the image. The reverse compensation is then conducted for each one accordingly.

3 Literature Review

Before carrying out appropriate compensation, the amount of energy attenuated within the range scene and luminance of the artificial light source are considered. In an UW image without artificial light, the background directly transmits natural light without reflecting off of objects and is therefore the brighter part of the image. Thus, if the luminance in the foreground of an image is higher than the background, it indicates the existence of an artificial light source. Figure 3.9 manifests some of the results by applying WCID.

Later on, inspired by [CC12], several approaches have been proposed which follow the same scope but apply changes to improve the result. One of the main concerns is to replace soft matting [HST11] with Median filter [Yan+11] or guided trigonometric bilateral filter [LLS13; SL14] to decrease the computation costs and number of iterations to speed up the running time.

Although the dark channel prior is widely used for UW image dehazing, but it may in fact fail to estimate the depth map due to the absorption property of water. The red color channel is absorbed as the first color, thus, dark channel cannot provide useful information when red color almost always has the lowest intensity between other color channels. Several attempts are made to replace the dark channel prior or to assist it with other alternatives [Li+16; LC16; GLW16; PC17].

Li et al. [Li+16] considered optical properties of UW images and treated each color channel separately. The algorithm estimates the global background light using the quad-three subdivision strategy [Kim+13]. Using this strategy, the image is divided into four equal rectangular regions, a score based on the average pixel value subtracted by the standard deviation of each region is then calculated. The region with the highest score is selected, within this region, the top 0.1% brightest pixels in the dark channel are chosen as candidates. The final global background light is the one with the maximum blue-red difference in the input image. This is done to remove the effects of bright objects in the foreground.

Rewriting the Equation 3.10 as

$$J_{\lambda}(x, y) = \frac{1}{t_{\lambda}(x)}(I_{\lambda}(x) - A_{\lambda}) + A_{\lambda}, \quad (3.14)$$

it is observed that the values of input image in the range determined by the transmission map are mapped to output values in the full dynamic range. This may cause underflow or overflow in those pixel values of input image which lie outside of this range. Thus, to avoid information loss, they considered the cost function

$$f_{loss} = \sum_{(x,y) \in \Omega} (\min(0, J_r(x, y))^2 + (\max(0, J_r(x, y) - 255))^2), \quad (3.15)$$

on the red channel and calculate the transmission map of this color channel accordingly. Once the transmission map of red channel is available, those that correspond to blue and green channels are calculated based on their relationship to the red channel [ZJQ15].

Having both the background light and transmission map of each color channel, the degraded UW image is restored. Furthermore, they employed an adaptive exposure map [TYW14] to fix the

probable introduced overexposure and underexposure of bright and dark regions of the restored image respectively. At the end, a contrast enhancement algorithm is also applied to improve the contrast and brightness of UW images.

Another alternative for dark channel prior is proposed by Peng and Cosman [PC17] where the depth map is estimated using image blurriness and light absorption. The motivation for this approach is to compensate for the shortcomings of the dark channel prior method especially in the presence of an artificial light. Moreover, the proposed method takes into account the situation when the background light is dim, which is ignored in the literature. For this, they combined three depth estimation methods to obtain the final scene depth.

The first depth map, $\tilde{d}_R(x, y)$, is obtained from the red channel where the points with higher red lights are assumed to be closer to the camera. This can take care of a scene with a relatively dim background since it can define the bright pixels as being close to the camera rather than being part of the background. The second depth map, $\tilde{d}_D(x, y)$, is estimated using the difference between the red channel to the maximum of the green and blue one, thus, the bigger value indicates the closer point to the camera. Using this map, one can take into account the images with a bright background. The third depth map, $\tilde{d}_B(x, y)$, is obtained based on image blurriness [PZC15], the pixels with higher blurriness are farther away from the camera. They used this map for the cases where the red light is almost completely absorbed so there is the need for another metric rather than colors to estimate the depth map. Finally, combining all three depth maps, they provide an accurate depth map of the image:

$$\tilde{d}_n(x, y) = \theta_b[\theta_a \tilde{d}_D(x, y) + (1 - \theta_a) \tilde{d}_R(x, y)] + (1 - \theta_b) \tilde{d}_B(x, y), \quad (3.16)$$

where

$$\begin{aligned} \theta_a &= S(\text{avg}(A_\lambda), 0.5), \quad \lambda \in r, g, b, \\ \theta_b &= S(\text{avg}(I_r), 0.1) \end{aligned} \quad (3.17)$$

are defined by the sigmoid function when $s = 32$, A_λ is the background light for each color channel λ , and I_r is the red color channel image,

$$S(a, v) = [1 + e^{-s(a-v)}]^{-1}. \quad (3.18)$$

The final depth map is further refined and smoothed by either soft matting [LLW08] or guided filtering [HST13].

Once the depth map is estimated, they approximated the transmission map $t(x, y)$ of the red color channel and calculated it for other color channels based on their relationships [ZJQ15]. The background light is estimated based on image blurriness and variance. They find three candidates with the highest blurriness and lowest variance from the top 0.1% blurry pixels in the input image, one candidate pixel for each RGB color channel.

Having the depth map and transmission map of the scene, the restored image is obtained using Equation 3.10. The results of this approach are evaluated qualitatively and quantitatively where, for quantitative comparisons, they simulated UW images using haze image formation (Equation

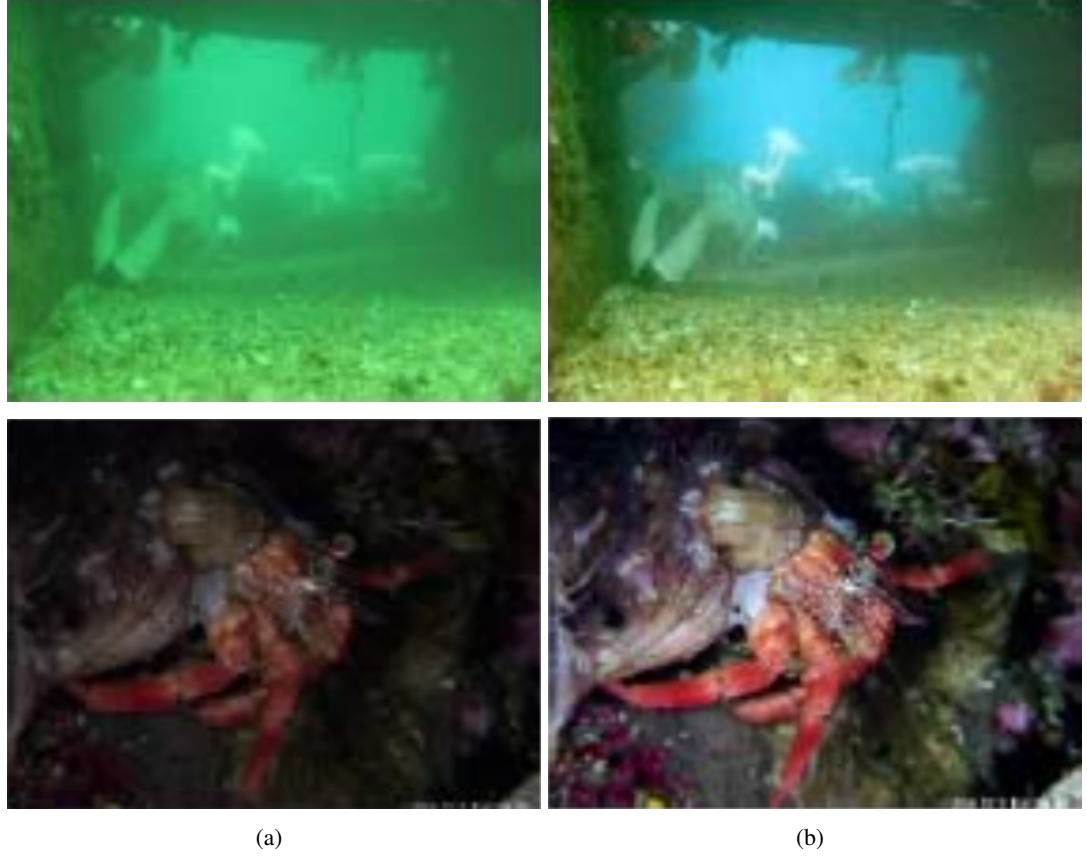


Figure 3.10: UW images courtesy of Peng and C.Cosman [PC17]. (a) Original image (b) dehazed result after applying the proposed method.

3.10) and adding blur PSF proportional to the known scene depth maps. PSNR and SSIM are used for numeric evaluation where the blind/reference image spatial quality evaluator (BRISQUE) [MMB12] and the UW image quality measure (UIQM) [PGA16b] are used for evaluation of non-reference real UW images. Figure 3.10 illustrates some of the results using this approach.

3.2.3 Other Approaches

Image restoration can be addressed without knowing the physical model directly. In this scenario, the implicit knowledge of the image data can be collected by training over a pair of data sets. This is possible when ground truth data or at least some features are available. The idea is that using training data sets, one can learn some statistical priors of the image class and using this information instead of an exact physical model, restore an approximation of the desired data. To this end, an approach towards color restoration is proposed by Torres-Méndez and Dudek [TD05].

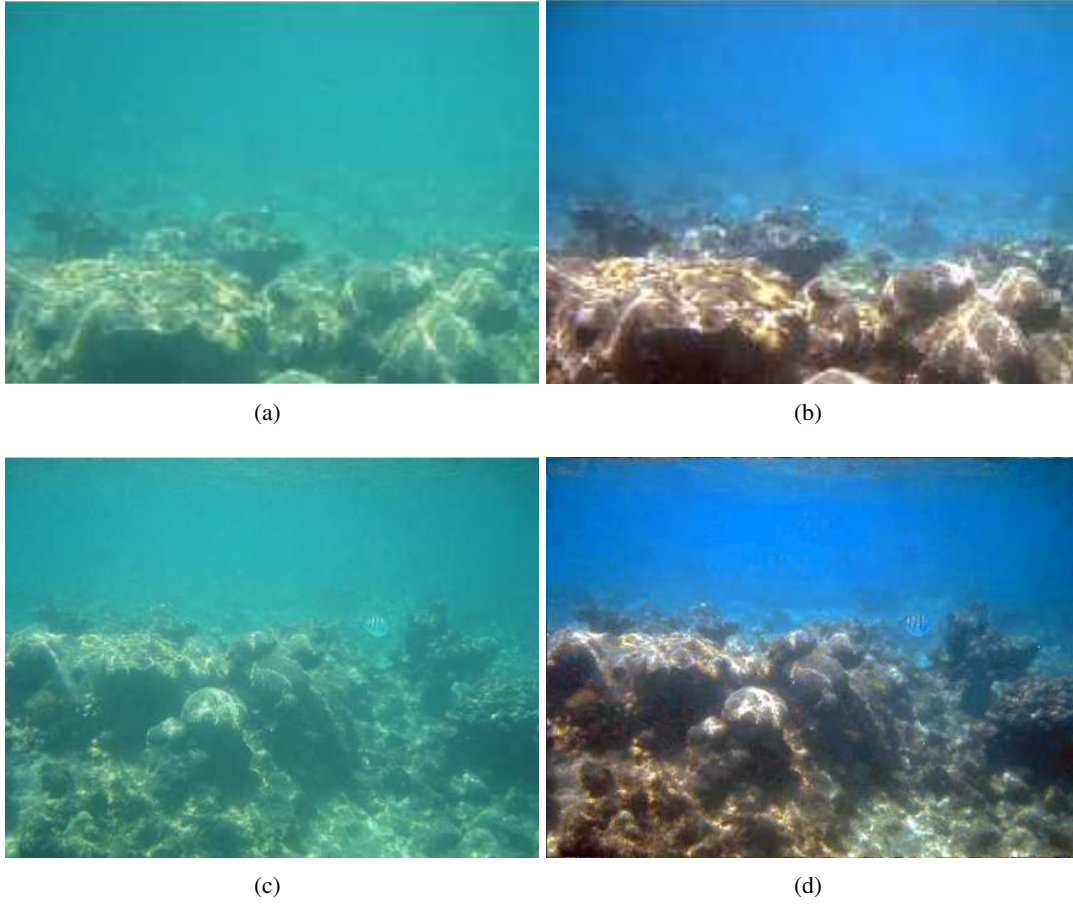


Figure 3.11: UW scenes from ocean environment. (a), (b) Training image pair captured at frame t . (c) Input image taken at frame $t + \delta$ and (d) the result after color correction. Image courtesy of Torres-Méndez and Dudek [TD05].

The approach is based on learning the statistical relationships from the training data using Markov Random Field (MRF). The training set consists of small patches of regions of interest that capture the maximum of intensity variations from the image to be restored. In order to preserve edges and prevent the results from over-smoothing colors, a high compatibility between neighborhood pixels is set for those which have similar colors and low compatibility between the one with a high change in color values.

The difference between colors should be measured in a way that matches perceptual similarity as good as possible. For this, they used the perceptual uniform color space CIE lab. Two scenarios are acknowledged. For the first one, they used ground truth data and degraded them using an attenuation filter. Some of the ground truth image patches are employed to conduct the training data. The second scenario is built upon the images which are taken by a robot in the sea. Once in a while the robot stops and takes a photo where an external light is used. This results in images



Figure 3.12: (a) Real UW image, (b) the result of super-resolution and descattering algorithm proposed by Lu et al. [Lu+17].

with a better color quality. These so-called high-quality images are the source of the training data to restore the neighboring depleted color frames. Figure (3.11) illustrates the result of this method.

Beside all degradations caused by light propagation UW, acquired images suffer from low-resolution due to using low-resolution cameras or difficulties in the transmission process which results in the compression of the data. Very few approaches have addressed this issue, though such examples are [Che+11; Lu+17].

Lu et al. [Lu+17] proposed a fusion based super-resolution and descattering approach. For this, first, they denoise and descatter the image simultaneously using non-local means (NLM) [BCM05] and a dark channel prior approach [Lu+15]. Next, both images noisy input image and the descattered and denoised image are super-resolved via a patch-based self-similarity algorithm proposed in [HSA15]. Finally, using a convex combination of two images from the last step, they recovered a high-resolution and descattered image. They used a fusion approach to combine the frequency, spatial and texture components of the image to blend the selective frequency bands and preserve the rich texture of local patches. Figure 3.12 illustrates an example of the results.

3.3 Underwater Image Enhancement

The second possible approach towards UW image degradations is to improve the quality of these images without knowing any prior knowledge such as attenuation coefficient, scattering and water type. These so-called enhancement methods are not based on a physical model of medium but make a total abstraction of the image formation process. Hence, they are usually faster and simpler than image restoration techniques. Color and contrast correction account for the most literature in this category.

Water absorbs the energy of light, therefore, going deeper in water, the more the light loses its energy and colors drop. Experiments show that the red color drops at the depth of approximately 3 m and is followed by orange, yellow, green and blue respectively. Blue color has the shortest wavelength/higher energy which allows it to travel deeper. As a consequence, a strong and non-uniform color cast and low contrast by means of depth characterize the UW images.

Here, the enhancement methods are summarized. Since these approaches do not follow any specific model, it is difficult to find the connection between them. Therefore, they are clustered in different categories based on their methodology.

3.3.1 Histogram and Contrast Stretching-based Approaches

One popular approach to improving color and contrast is to apply different variations of color equalization, histogram equalization, and contrast stretching.

A piece of pioneering work has been done by Chambah et al. [Cha+03] based on ACE [RGM03] to compensate the strong non-uniform color cast due to the depth of water and artificial illumination. This was done to achieve a better segmentation and recognition of fish species. They employed diverse images taken in a fish tank and balanced the color cast before segmentation. ACE represents Automatic Color Equalization which is inspired by some adaptation mechanisms of the human visual system. It merges the grey world hypothesis and white patch equalization mechanisms while taking into account the spatial distribution of color information.

Later, Iqbal et al. [Iqb+10] submitted a method concerning color and contrast correction which is based on color balancing and contrast correction in RGB and HSI color spaces. This is a follow-up of the work that they did earlier [Iqb+07]. First of all, the histogram of RGB channels are stretched towards the upper side, lower side or both sides based on their values. The channel with the lowest value is stretched towards the maximum intensity and the channel with the highest value, towards the minimum intensity and consequently, the channel with the value in between, is stretched towards both sides. In order to avoid the problems with single outlying pixels having too high or too low a value, they discarded 0.2% most extreme values on each side. Finally, they applied contrast correction to both sides of Saturation and Intensity parameters of HSI color space to provide a wider color range and recovered the true color of UW images. Edge detection and histogram representation are used to evaluate the performance of the proposed method. Figure 3.13 illustrates some of the results before and after applying this approach.

Inspired by the method proposed by Iqbal et al. [Iqb+10], Ghani and Isa [GI14] presented an approach to improve the color and contrast while enhancing the details. It is a two stage method, where contrast correction is the main contribution of the first stage and color correction is followed as the second stage. With the first step, a global histogram stretching is applied on each RGB channel, then the stretched histograms are divided into lower and upper sides based on their average value. Each side is then stretched via Rayleigh distribution to the entire dynamic range.

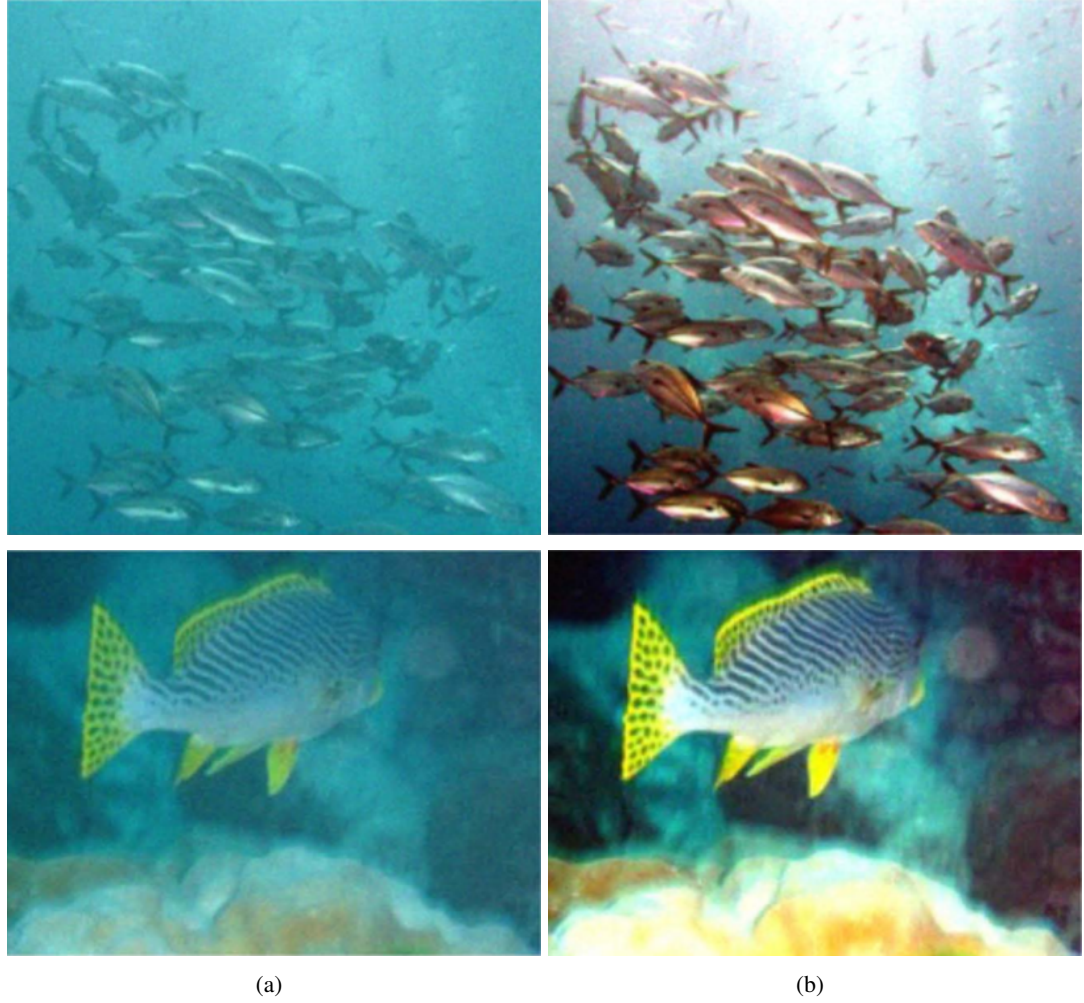


Figure 3.13: UW scenes from natural environment. Original image (a) and color corrected image (b). Image courtesy of Iqbal et al. [Iqb+10].

As a result, two images are produced from the original one, one under-saturated and one over-saturated image from the lower and upper sides of the histogram respectively. Next, these two RGB images are combined by means of an average value. At the second step, the output from the last step is converted to HSV color space and the S and the V components are stretched to enhance the colors. They evaluated their approach's performance by examining an image entropy for image details, mean square error (MSE) and signal to noise ratio (PSNR) for amount of noise.

Originally, contrast limited adaptive histogram equalization (CLAHE) was developed for medical images [Pis+98]. However, despite the good results, it introduces noise and unnatural brightness. To reduce the artifacts, Hitam et al. [Hit+13; Yus+13] proposed mixed CLAHE by combining the results of this method on RGB and similarly HSV color spaces. They applied

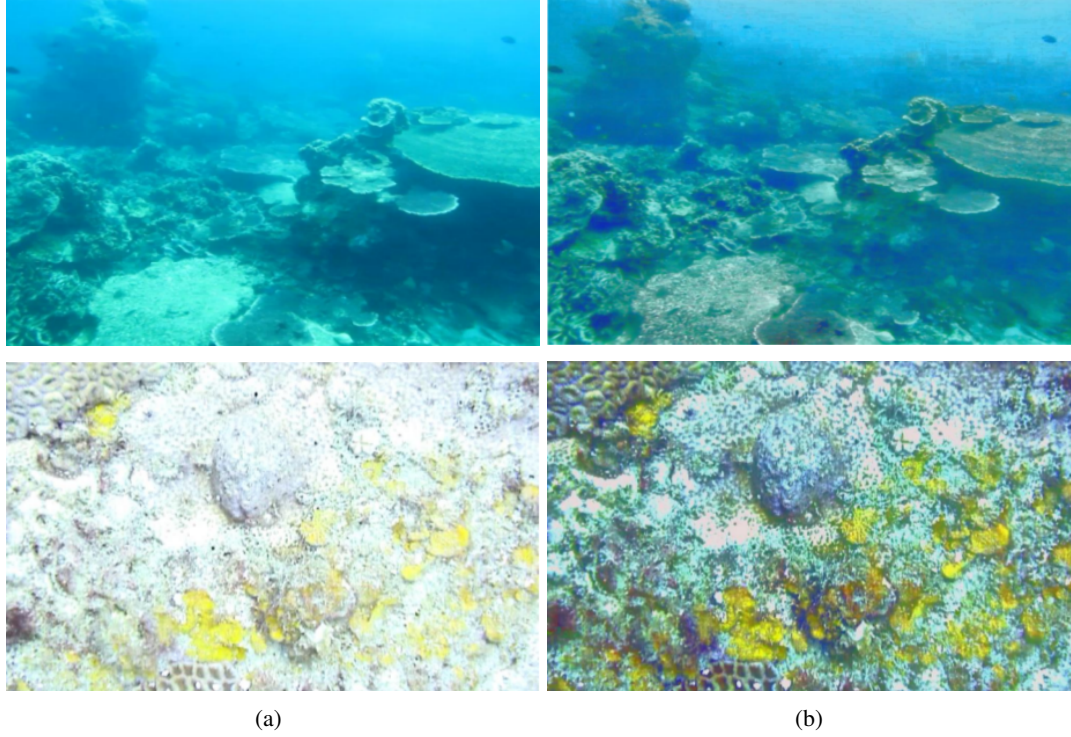


Figure 3.14: UW scenes from Redang island, Malaysia. (a) Original image and (b) the enhanced result. Image courtesy of Hitam et al. [Hit+13].

CLAHE to all three RGB channels [Hit+13] or applying it only to R channel of RGB color space [Yus+13] since the loss of red color is more critical than green and blue. Next, this result is normalized and integrated with the result of applying CLAHE to the V and the S channels of HSV using Euclidean norm. Some results are provided to evaluate the performance of this method (Figure 3.14).

Considering UW image distortions, one of the degradation sources which has a significant impact on the performance of image processing algorithms, that of floating particles, has been neglected. These particles consist of dead material and dissolved organic (so-called "marine snow") and inorganic matters which not only disturb the scene visibility but degrade the performance of computer vision tasks.

Boffety and Galland [BG12] proposed a phenomenological model of marine snow to use this degradation source as a priori knowledge to improve the results of image enhancement algorithms. The effect of such a phenomena on one possible application, color correction, has been investigated and ultimately, were able to provide a solution to improve the robustness of such an algorithm. To this end, they first simulated marine snow where the water effect is also considered. It is integrated into the backscatter component proposed by McGlamery [McG80]. The backscatter component $E_{b,\lambda}(x,y)$ for each color channel ($\lambda \in (Red, Green, Blue)$)

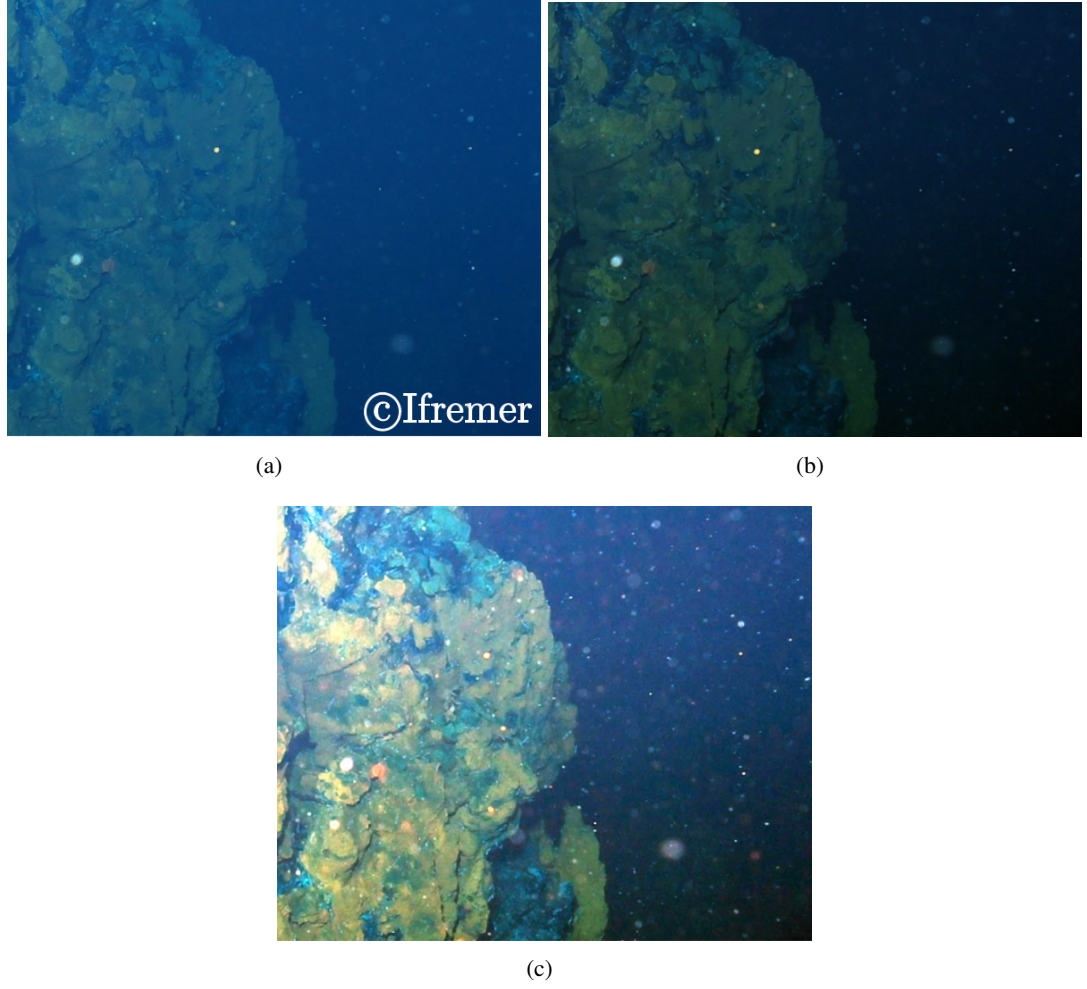


Figure 3.15: UW scene in natural environment (provided by Ifremer). From left to right: original image, Result after applying HSM [Iqb+07], result after considering the effect of marine snow. Image courtesy of Boffety and Galland [BG12].

is calculated via summation of the scatter contribution due to each volume of water $E_{bk,\lambda}$ (where water column is divided into thin slabs) and the snow content in that volume (if any) $E_{bk,\lambda}^{snow}$ as follows:

$$E_{b,\lambda}(x,y) = \sum_{k=1}^{Z_0/\Delta z} (E_{bk,\lambda} + E_{bk,\lambda}^{snow}), \quad (3.19)$$

here Δz is the height of each volume, Z_0 is its distance to the detector. Marine snow is considered to behave as white Lambertian scatter with a Gaussian function as the reflection coefficient. Thus, white Gaussian-shaped particles are positioned based on salt & pepper distribution, considering in each volume of slab only one particle can be found.

Once the model is completed, they simulated UW images with and without marine snow considering two different water types (clear and turbid). To analyse the effects of macro-particles on image processing algorithms, they applied the histogram stretching method (HSM) [Iqb+07] on all simulated data. The method failed in the presence of marine snow, though it gave interesting results in the case of color correction without marine snow. The hypothesis is derived that the bright spots of marine snow limit the dynamic range of the color adjustment. As a solution, to enhance the robustness of this algorithm where marine snow is present, they replaced the maximum value in the calculation by 99% quantile value. This is done since the high-intensity values of the image seem to correspond to the marine snow bright spots. The outcome of such a reformulation is shown in Figure (3.15).

Khan et al. [Kha+16] proposed a wavelet based image enhancement which compensates color and contrast loss of hazy UW images. First, from a single degraded input image, they conduct two intermediate images. The first intermediate image is obtained by applying histogram equalization on V channel of image in HSV color space and afterward, on all three RGB color channels. This image is a color corrected version of the degraded input image. They obtained the second intermediate image by applying a modified version of CLAHE algorithm. This is done by redistributing the clipping limit among all the histogram bins in an iterative manner until the excessive area is negligible. This image is a contrast corrected version of the input image.

Once both intermediate images are available, they applied two levels of wavelet transform on both images and fused their coefficients by means of maximum values. The final enhanced image is acquired by taking an inverse transform of the fused coefficients.

3.3.2 Filter-based Enhancement Approaches

This category contains the techniques which enhance the general quality of UW images by applying a sequence of filters such as [AMK05a; Baz+06; PK12].

Arnold-Bos et al. [AMK05a] proposed a denoising method as a pre-processing step for edge detection. The approach consists of several filters, each addressing one degradation. Non-uniform lighting caused by backscattering is considered as the first noise which is reduced by applying a contrast equalization. Assuming $I(x, y)$ as an original image and $I_{LP}(x, y)$ as its low-pass filtered version, the scaled contrast equalized version is:

$$I_{eq}(x, y) = I(x, y) / I_{LP}(x, y). \quad (3.20)$$

This step is followed by histogram clipping to get rid of outliers. The remaining noise sources such as sensor noise and floating particles are suppressed using an adaptive smoothing filtering to preserve the good visibility in the presence of edges. Two classes of adaptive smoothing methods are used, anisotropic filtering [PM90] and Kovsi's phase preserving wavelet filtering [Kov99]. Both improved the edge detection although the wavelet-based method enjoys the self-thresholding mechanism which allows for a fully automatic approach. The quality of this method is evaluated by gradient magnitude histogram. In addition, they suggested a simple condensed numerical

3 Literature Review

criterion as follows:

$$R = \exp\left(-\frac{1}{N' - 1} \sum_{i=1}^{N'} (\ln h(i) - \ln \bar{h}(i))^2\right), \quad (3.21)$$

where i is the gradient magnitude (its range is considered from $i = 1$ to $N' = 2/3N$ where typically $N = 128$), and $h(i)$ and $\ln \bar{h}(i)$ are the gradient magnitude histogram and the result after applying a linear regression on $\ln h(i)$ respectively. This criterion gives a grade from zero to one to an image which is based on similarity of the distribution of the gradient magnitude histogram to an exponential distribution. The closer the robustness is to one, the better the enhancement is, i.e. a well-contrasted and noise free image is marked as one.

The aforementioned method is proposed to denoise the grayscale images so the color cast is not considered to be compensated. To cover the color cast besides the other noises, Bazeille et al. [Baz+06] presented a pre-processing algorithm as the preliminary step of a feature point extraction and edge detection. The algorithm contains several independent steps towards correcting non-uniform lightening, noise reduction and color and contrast adjustment.

In order to sharpen the edges, the low frequencies (illumination) are decreased and mid and high frequencies (reflectance) are amplified. This is known as homomorphic filtering. This step is followed by a wavelet denoising algorithm to suppress the Gaussian noise and later an anisotropic filter to smooth the homogeneous area for better segmentation. In the end, contrast stretching and color equalization are applied to provide a more pleasant image. They applied the approach on natural UW images with and without synthetic UW degradations. They simulated the degradations using the Jaffe-McGlamery model for blur and non-uniform illumination, Gaussian and particle noise as an additive contribution and finally, they reduced the color range by histogram operation. The overall performance of the method is evaluated by gradient magnitude histogram and a robustness criterion proposed in [AMK05a]. Some of the results are shown in Figure (3.16).

Zhang et al. [Zha+17] proposed an image enhancement method inspired by the Retinex framework. The Retinex theory proposed by Land et al. [LM71] simulates the human vision system and attempts to achieve the color constancy and discounts changes in illumination. The work proposed by Shu Zhang et al. is based on multi-scale Retinex (MSR) [JRW97] where usually, three scales (small, intermediate and large) are used to handle different details in the image. The reason for such an approach is the spatially variant luminance in both Retinex and UW situation. The approach employs MSR algorithm in LAB color space. Unlike traditional Retinex which uses a Gaussian filter and only considers variations in the spatial domain, they employed a combination of bilateral [TM98] and trilateral [CT05] filters for different color channels. Thus, degradations of UW images present in different domains are taken care of. After converting the RGB image to LAB color space, they applied a bilateral filter to the L channel and a trilateral filter to A and B channels. This is when the bilateral filter takes into account both uneven distortion in the spatial domain and intensity values, and the trilateral filter analyzes the distribution of the intensity gradients within the filter range. Therefore, it can suppress the halo artifacts caused by MSR.

3.3 Underwater Image Enhancement



(a) Without additional noise



(b) With additional noise

Figure 3.16: UW scenes extracted from TOPVISION videos. The performance of method is evaluated in two cases: without additional noise (a), and in presence of additional noise (b). Original image (first column in both (a) and (b)) and result after enhancement (second column in both (a) and (b)). Image courtesy of Bazeille et al. [Baz+06].

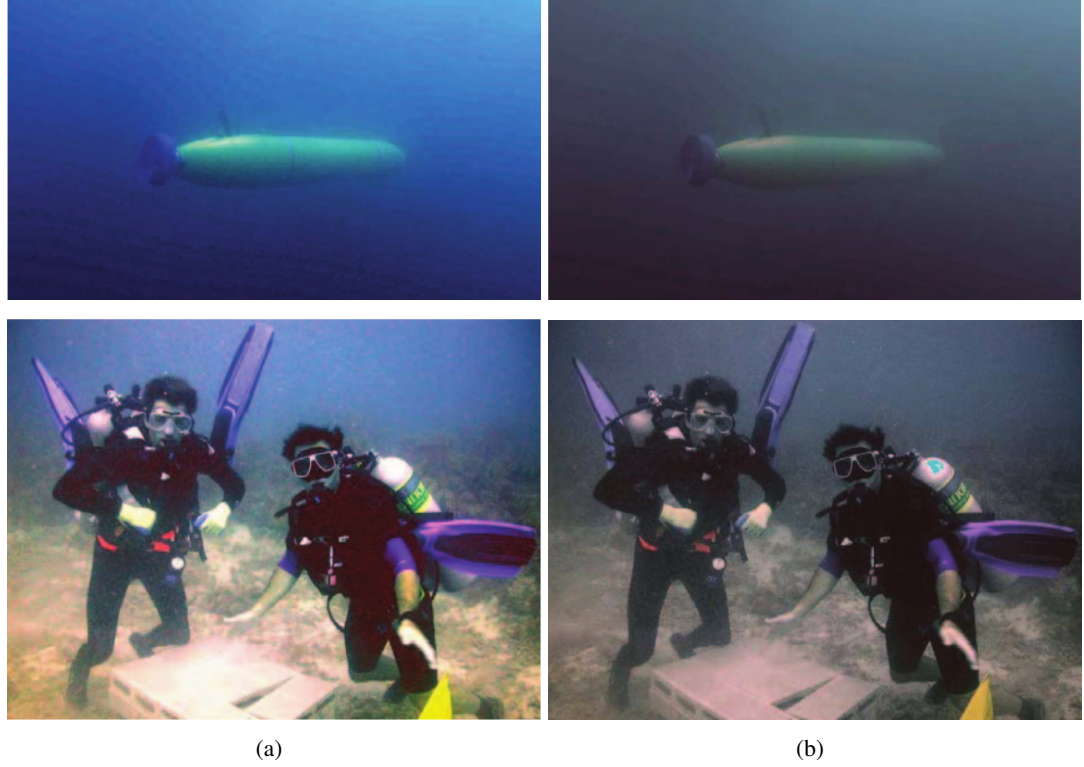


Figure 3.17: UW scenes in natural environment. (a) Original image and (b) result of quaternionic image correction method. Image courtesy of Petit et al. [PCC09].

Furthermore, they simulate the scattering and absorption effect by adding milk and grape juice in clear water respectively. Thus, using target images, they provide several test images with different turbidity conditions and evaluate the performance of their algorithm accordingly.

3.3.3 Other Approaches

A very different and interesting approach is reported by Petit et al. [PCC09]. They used quaternion transformation and separated the water pixels from the rest of the image and returned those towards the grey to achieve color and contrast correction. Quaternions (or hyper-complex numbers) are an extension of complex numbers to four dimensions, proposed by Hamilton 1843. By this, reflection, rotation and scaling can be expressed in quaternion space. (For more on this topic, the reader is referred to [San02]). Using Beer-Lambert law, the hue vector associated with the water color is calculated. Once the vectors are obtained, the water pixels are moved close to the hue axis (near the gray axis), though the color of objects are preserved. Therefore, they compensated the attenuation effect on objects in RGB color space. Additionally, they showed that the hue vector can be approximated via Principle Component Analysis (PCA) instead of using Beer-Lambert law and therefore, avoid the expensive computation costs. In Figure (3.17) two

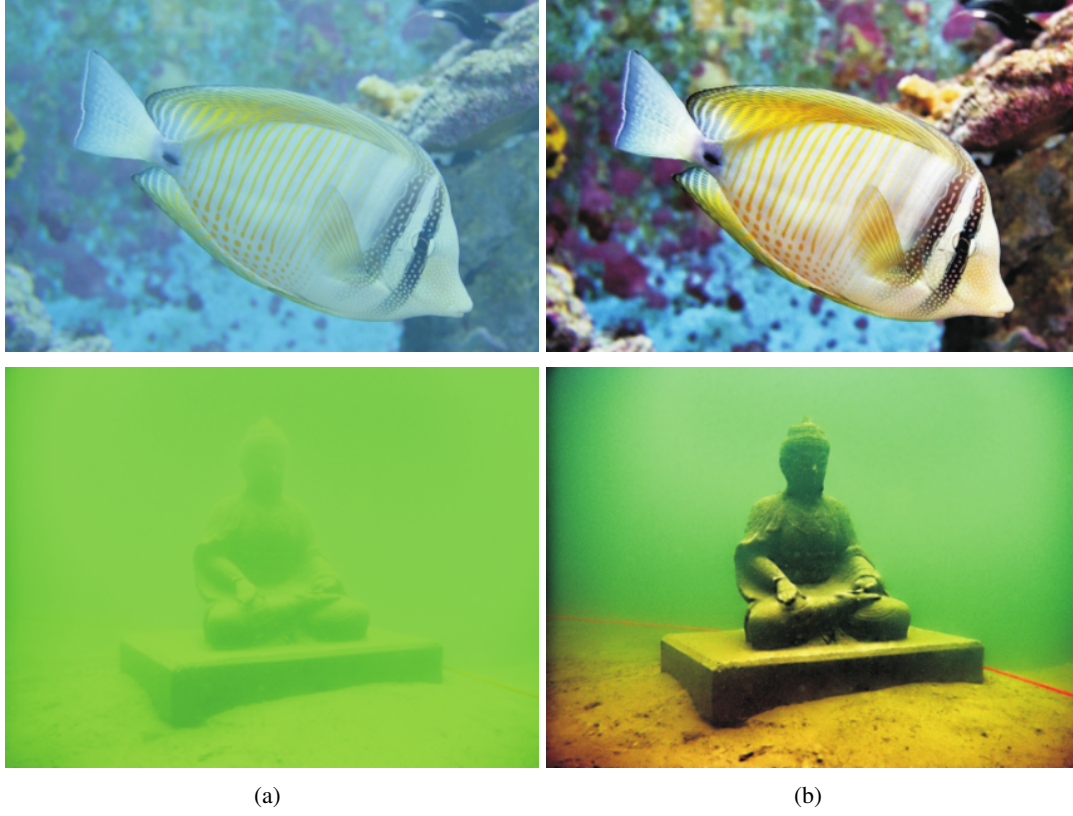


Figure 3.18: UW scenes from natural environment. Original image (a) and the result after enhancement (b). Image courtesy of Ancuti et al. [Anc+12].

pairs of images are shown before and after applying the outlined method.

A new perspective in UW image enhancement is introduced [Anc+12; Fan+13] which is based on a fusion strategy and can be designed to enhance color and contrast, reduce the noise and provide better details. Ancuti et al. [Anc+12] proposed such a method to reduce the noise, enhance the global contrast, details, and edges while it improves the exposedness of dark regions. The approach is based on two inputs derived from the single degraded image and four weight maps which aim to increase the visibility of the distance objects.

First, the degraded image is white balanced to remove the color cast and marked as the first input image. This input is denoised and the contrast local adaptive histogram equalization [Zim+88] is applied in order to obtain the second input. They chose the corresponding weights to give a better color appearance, enhance the global and local contrast and improve the salient features and exposedness accordingly. In order to avoid undesired halo artifacts, they applied the method in a multi-scale fashion. Each input is decomposed into a Laplacian pyramid and similarly, for each normalized weight map, a Gaussian pyramid is computed. Finally, they obtained the fused pyramid by combining both pyramids. The method is evaluated in case of dif-

ferent applications such as local feature points matching, segmentation and image dehazing. Two examples of test images before and after applying Ancuti et al. method are shown in Figure (3.18).

3.4 Underwater Image Denoising

Observed images are often corrupted by acquisition channel or artificial editing. Therefore, image denoising is an important image processing algorithm. The goal of image denoising procedure is to map from a noisy image to an output image where the noise has been reduced. The main properties of a good image denoising model are that it removes noise while preserving edges which is useful by itself and an important pre-processing step for many other applications.

Although image denoising is a common restoration problem and could be categorized in the image restoration group, here it is mentioned in a separate section. This is due to the existence of a special unwanted signal (suspended particles) in UW images which need to be addressed from a different point of view than common imaging noises. From the perspective of image processing and computer vision tasks, floating particles (commonly known as marine snow) are considered as noise, however, they are objects in the scene. This is due to the difficulties that these particles cause, therefore, here denoising algorithms are summarized specifically to evaluate their performance in eliminating this phenomenon.

There are many and diverse approaches which are specifically introduced to reduce noise from an UW image such as [PK10; FXG11; JG12; Liu+15a]. Following, a short summary of some of these approaches is provided.

It is common practice to use various multi-resolution transforms including wavelet, curvelet, and contourlet for image denoising. This is due to their ability to extract the image features, low entropy, and multi-resolution characteristics. Prabhakar and Praveen [PK10] introduced an adaptive wavelet subband thresholding algorithm to reduce additive noise from UW images. Floating particles (marine snow) have been considered as unwanted signals of the size of one pixel which are introduced as an additive noise of distribution that is reasonably similar to Gaussian. For this, they applied homomorphic and then anisotropic filtering to correct non-uniform illumination of light and further smooth the homogeneous area while preserving image features respectively. At the end, they applied wavelet transform and thresholded the small noisy coefficients out to reduce the noise while reconstructing the essential signal characteristics. They used a modified BayesShrink function to define a threshold for each subband accordingly. The results are compared to different wavelet shrinkage functions such as BayesShrink [CYV00], NormalShrink [KGC02], and adaptive subband thresholding [SSS07].

One important aspect of image denoising is to preserve the image details and edges as much as possible while reducing the noise in both high and low frequencies. To this end, there are approaches which combine 'let' family transforms with high-pass filtering [FXG11; SSS13] or Adaptive Total Variation [JG12] for better denoising results.

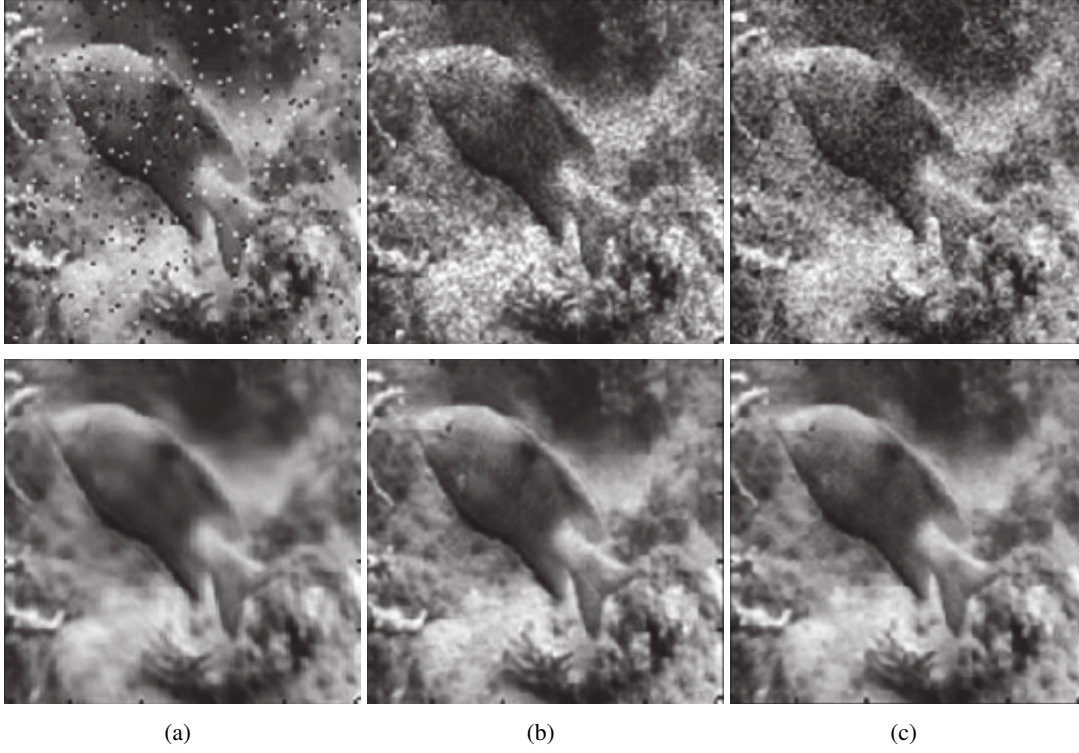


Figure 3.19: UW scene from natural environment corrupted by salt & pepper (a), speckle (b) and Gaussian (c) noise. First row is the original image and second row is the result after denoising. Image courtesy of Shanmugsundaram et al. [SSS13].

Feifei et al. [FXG11] considered that UW noise is mainly due to backscatter light which performs as the color noise with significant low-frequency components. To prove this, they cut a small area of an UW image where there is no object and the light intensity is uniform. Then, they analyzed the power spectrum density of this area. Their conclusion was that an UW image is mainly corrupted with color noise which its energy is concentrated in the low-frequency area. Thus, to suppress the low-frequency components of backscatter noise as well as uncorrelated high frequency noise, they introduced an approach which combines wavelet decomposition with high-pass filtering. After wavelet decomposition, the low frequency-component is filtered using Butterworth high-pass filter when the high frequency-component is thresholded using the VisuShrink function.

Later, Shanmugsundaram et al. [SSS13] applied the same idea of combining "let" family with a high pass-filter. They presented an UW image denoising algorithm by fusing two intermediate images of the same. For this, a Gaussian filter together with a curvelet transform is used. They claimed that UW images are polluted by three basic noises, namely, Gaussian, speckle, and salt & pepper noise. To suppress these kinds of noises, they implemented a Gaussian filter

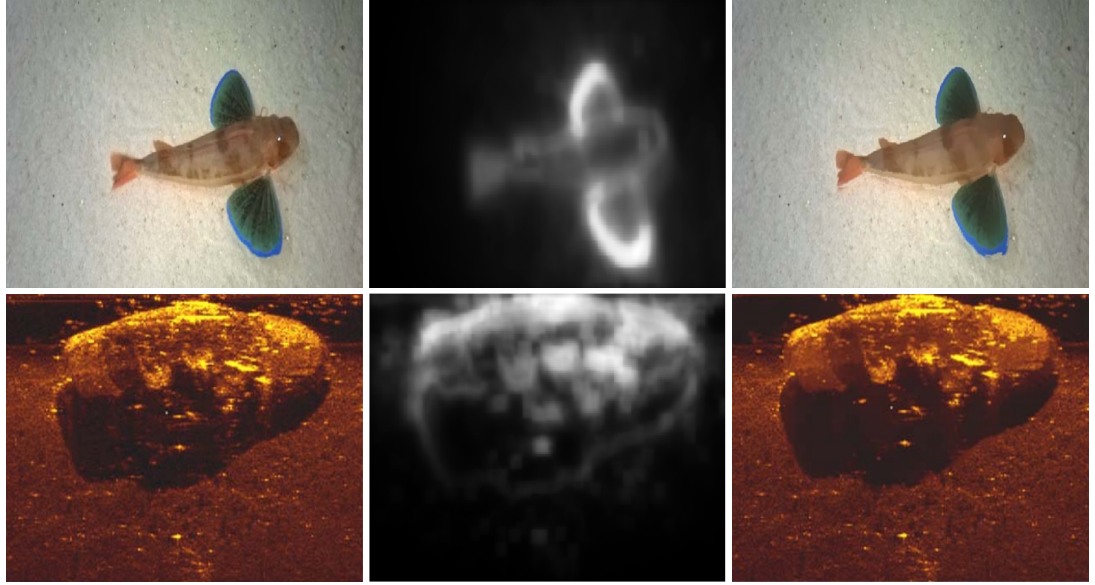


Figure 3.20: From left to right: original image, saliency map and the result after denoising. Image courtesy of Liu et al. [Liu+15a].

using Euclidean transformation and replaced each pixel value by the average distance of every neighborhood pixel. Consequently, the first intermediate image is a smoothed version of the original image using a Gaussian kernel as follows:

$$I_1(x, y) = I(x, y) \cdot \exp\left(\frac{-x^2 - y^2}{2\sigma^2}\right) / 2\pi\sigma^2, \quad (3.22)$$

where $I(x, y)$ and $I_1(x, y)$ are the original and the image after Gaussian filtering respectively. A second intermediate image $I_2(x, y)$ is obtained by denoising the original image through curvelet transform. After decomposition, the coefficients are thresholded to reduce the noise. The final image is conducted by averaging I_1 and I_2 . The results of this denoising engine are shown in Figure 3.19. Although the results display a better quality compared to each intermediate image alone, they nevertheless, still suffer from blurring image details.

A second strategy is filtering the noisy images in the spatial domain such as [Liu+15a; Ban+14]. Liu et al. [Liu+15a] introduced an UW image denoising based on weighted Median filtering and saliency map. Saliency detection is used to speed up the process by detecting and prominent the salient regions. Using a patch based approach, they achieved a pixel accurate saliency map which covers the objects of interest and consistently separates fore- and background. They employed weighted Median filtering to avoid strong blurring of the edges. For this, they applied the median tracking strategy to seek median in the joint histogram. This reduces computation costs since the chance of having similar pixels in a neighborhood is very high. This is done by shifting the center pixel of the local window to the left or right according to the positive or negative balance of weighted Median filter of adjacent windows. Thus they achieved a sparse approach which speeds the procedure while suppressing noise. Some of the results of this method are shown in

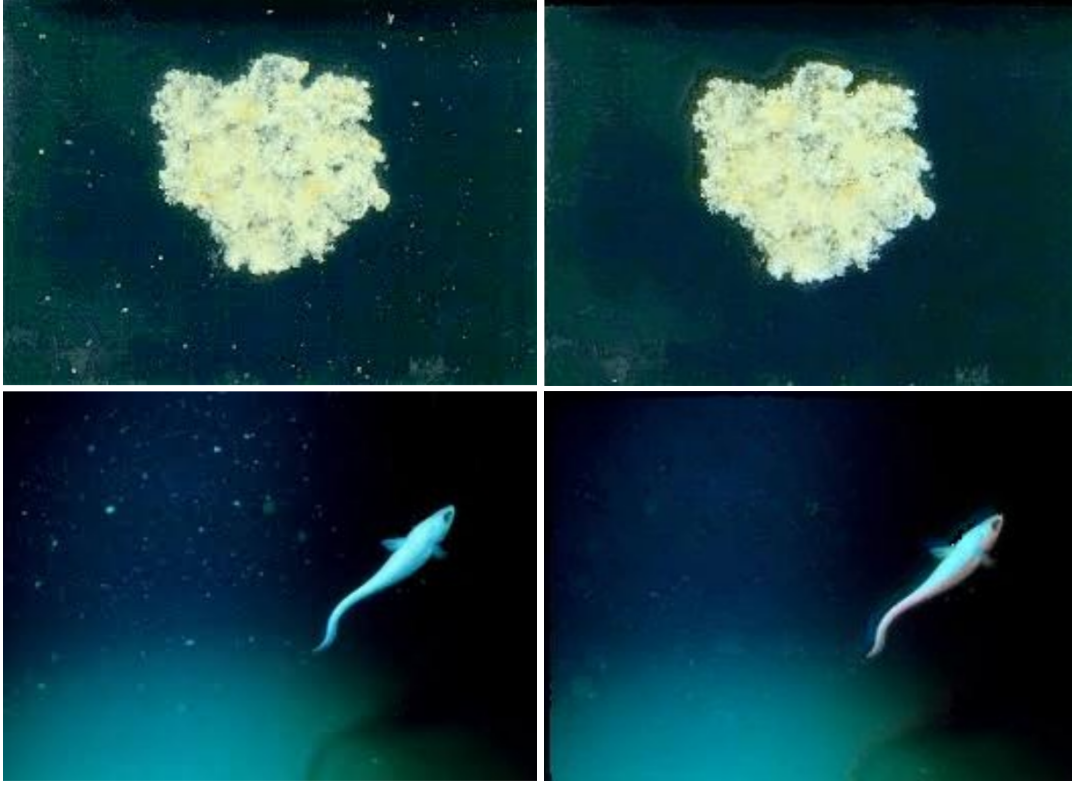


Figure 3.21: From left to right: original image and the result after denoising. Image courtesy of Banerjee et al. [Ban+14].

Figure 3.20. Despite the noise reduction, the details are smoothed, e.g. they lost the fish details in the first image.

Unlike the aforementioned methods which considered any kind of UW noises as one of the additive noises such as Gaussian, color, and salt & pepper, Banerjee et al. [Ban+14] proposed an adaptive probabilistic approach to eliminate the floating particles, marine snow effect, from UW images. They defined this unwanted signal as tiny sparkling dots consisting of two to three pixels which reduce the scene perception. To eliminate this effect, they sliced the luminance channel of YCbCr color space in square windows of size 7×7 . The slices with a high variance are considered to be the most probable locations of having snowy dots. Therefore, they calculated the probability of having snowy dots in a window as follows:

$$P(MS) = 1 - \frac{N_{HL}}{N}, \quad (3.23)$$

here, N_{HL} and N are the number of a high luminance pixels and the total number of pixels in a local window respectively. The idea is that the less the probability of high luminance pixel, the more the probability of having snowy dots in a local window. Misclassification may occur in a case where the window is on the edge or boundary of an object with high luminance. To avoid

this, they double checked $P(MS)$ by stretching the window two pixels from both sides without changing the center. If the number of high luminance pixels is increased, it denotes a part of a true object and $P(MS)$ is set to a very low value. Otherwise, the window is considered as being corrupted with snowy dots, therefore, the center pixel is replaced by the median value of the local window. Figure 3.21 illustrates some of the results.

3.5 Summary

The reviewed approaches are classified based on their methodology, main objective and application (Table 3.1 and 3.2). The tables correspond to the image restoration and image enhancement techniques respectively.

Once the user has control over the data acquisition, using compact hardware and proper light source introduced in [SK05; TS09] the visibility of UW images as regards dehazing and color correction can be enhanced significantly. In respect of blur UW, the PSF of a water body can be obtained using an *in situ* calculation and employing Wiener filtering which can then restore blurry UW images in turbid water. However, mostly, the low visibility and degraded images are given to the user thus, the approach for recovering a high quality image is limited to software methods. In view of the table 3.1 and 3.2, one can observe that dehazing and deblurring are mainly addressed via image restoration methodology where color correction is the main objective for image enhancement approaches.

For image dehazing, most of the approaches consider methods such as dark channel to estimate the background light and, using this information, remove the veiling light. color correction is mostly part of the procedure in these approaches. Among all, Chiang and Chen [CC12] provide a comprehensive approach for dehazing where the camera-object distance, depth of scene and range of scene are taken into account to remove haze precisely. This approach is suggested if the images are taken where artificial light is present. The dehazing approach proposed by Bianco [CME10] provides very successful color correction together with dehazing. It especially achieves good results for recovering true colors of the objects at distance and defining the borders of the objects with water. Despite its successful color correction and dehazing, it introduces over-exposure and leads to losing some of the image details. In a situation where the depth map cannot be estimated accurately due to the lighting condition, the approach proposed by Peng and C.Cosman [PC17] or Li et al. [Li+16] are suggested. In these approaches, authors considered blurriness and light absorption, and dark channel and light channel to estimate the respective depth map. This reduces the error in estimating background light and, consequently, the transmission map which results in a more accurate dehazing approach.

Table 3.1: Summary of UW Image Restoration Algorithms. The sign \checkmark defines the main objectives of the methods where \star defines the by-product achievements.

Software	Hardware		color correction	Dehazing	Contrast enh.	Deblurring	Denoising	Super-resolution
\checkmark	\checkmark	Zhishen et al. [ZTG03] 2003				\checkmark		
\checkmark	\checkmark	Schechner and Karpel[SK05] 2005	\checkmark	\checkmark	\checkmark			
\checkmark	\checkmark	Treibitz and Schechner[TS09] 2009	\checkmark	\checkmark				
\checkmark		Torres et al. [TD05] 2005	\checkmark					
\checkmark		Trucco and Olmos[TO06] 2006	\star			\checkmark		
\checkmark		Hou et al. [Hou+07a] 2007			\checkmark		\checkmark	
\checkmark		Bianco et al. [CME10] 2010	\star	\checkmark				
\checkmark		Wang et al. [WZF11] 2011		\checkmark				
\checkmark		Feifei et al. [FXG11] 2011					\checkmark	
\checkmark		Chiang and Chen [CC12] 2012	\checkmark	\checkmark				
\checkmark		Wu and Li [WL13] 2013			\star	\checkmark		
\checkmark		Shanmugsundaram et al. [SSS13] 2013					\checkmark	
\checkmark		Stephan and Beyerer [SB14] 2014		\checkmark		\checkmark		
\checkmark		Banerjee et al. [Ban+14] 2014					\checkmark	
\checkmark		Liu et al. [Liu+15a] 2015					\checkmark	
\checkmark		Li et al. [Li+16] 2016	\checkmark	\checkmark	\checkmark			
\checkmark		Peng and C.Cosman [PC17] 2017	\checkmark	\checkmark		\star		
\checkmark		Lu et al. [Lu+17] 2017		\checkmark			\checkmark	\checkmark
\checkmark		Farhadifard et al. [FZL15] 2015	\checkmark			\checkmark		\checkmark
\checkmark		Farhadifard et al. [FR16] 2016			\star	\checkmark		
\checkmark		Farhadifard et al. [FRL17b] 2017					\checkmark	
\checkmark		Farhadifard et al. [FRL17a] 2017					\checkmark	

For image deblurring, two approaches specifically address the issue via inverse filtering [TO06; WL13]. Both are simple and effective where [TO06] is faster and improves the color cast. Although the approach proposed by Wu and Li [WL13] is only applied on UW images taken in an artificial pool, the results show the success of the algorithm at deblurring where some image details, hidden due to slight backscattering, are recovered successfully.

3 Literature Review

Table 3.2: Summary of UW Image Enhancement Algorithms. The sign \checkmark defines the main objectives of the methods where \star defines the by-product achievements.

	color correction	Dehazing	Contrast enh.	Deblurring	Denoising
Chambah et al. [Cha+03] 2003	\checkmark				
Arnold-Bos et al. [AMK05a] 2005			\checkmark		\checkmark
Bazeille et al. [Baz+06] 2006	\checkmark		\checkmark	\star	\checkmark
Iqbal et al. [Iqb+07] 2007, [Iqb+10] 2010	\checkmark		\star		
Petit et al. [PCC09] 2009	\checkmark				
Boffety et al. [BG12] 2012	\checkmark				
Ancuti et al. [Anc+12] 2012	\checkmark		\checkmark		\checkmark
Hitam et al. [Hit+13], [Yus+13] 2013	\checkmark		\checkmark		
Ghani and Isa [GI14] 2014	\checkmark		\checkmark		
Khan et al. [Kha+16] 2016	\checkmark	\checkmark	\checkmark		
Zhang et al. [Zha+17] 2017	\checkmark	\checkmark	\checkmark		

color correction and contrast enhancement are mainly addressed via image enhancement algorithms. The main aim is to provide images with a visually better color quality and contrast. This is where there is no reference data for evaluating the color accuracy. These approaches are mostly based on contrast and histogram equalization which provide images with a full dynamic range leading to color and contrast improvement. They are mostly faster and simpler than image restoration approaches and do not need any specific parameters. Among these approaches, fusion based methods such as [Anc+12; Kha+16] can handle more than one objective at a time since multiple enhanced versions of degraded input image conduct the final result. Each enhanced version tackles one of the degradations, thus, the fused version is the enhanced version of an input image in two or more aspects. Due to the fact that these methods do not follow a specific physical model, it is difficult to score them compared to each other. All of them provide visual improvement however, depending on the data it may vary from one algorithm to another.

UW image denoising approaches mostly considered the additive noises together with floating particles in one category and proposed methods such as wavelet thresholding, curvelet or high-pass filtering. Despite their success at removing noises such as Gaussian, salt & pepper and pickle, they cannot remove floating particles with sizes bigger than a single pixel. One of few approaches capable of removing floating particles is proposed by Banerjee et al. [Ban+14].

Low-resolution of UW images is mainly due to using low-resolution camera or image compression which is applied to ease the process of data transmission. Although it is more a hardware shortcoming rather than a medium effect, it remains, however, a common problem for UW images and still needs to be addressed. Very few approaches have tackled the problem by applying the

general super-resolution algorithms for UW data i.e. Chen et.al [Che+11; Lu+17].

3.6 Conclusion

In this chapter, the relevant methods towards UW image restoration, enhancement, and denoising have been reviewed. Although image denoising could be categorized in image restoration group, but it is mentioned separately. This is due to the presence of a special unwanted signal (floating particles) in UW images which needs to be addressed from a different point of view than those of common imaging noises.

After reviewing the existing literature, a section was provided which classifies the approaches based on their methodology, main objective and application. The best choices for each of the UW degradations, color cast, haze, blur, and noise are introduced. This could give the reader a compact and comprehensive overview of the challenges and their solutions.

In view of this survey, the following conclusions have been made which illustrate the open issues:

- Among the UW image degradations, color cast and haze are addressed often, resulting in some rather rich literature in these two topics. However, limited literature has addressed blur since it is not considered to be as visually disturbing as color cast and haze. Yet it degrades image quality especially at the image structures, leading to poor performance of advance computer vision tasks i.e. segmentation.
- Despite the fact that image transmission is still a challenge for UW data, there is hardly any literature on UW image super-resolution.
- Despite the important effect of suspended particles in image degradations and application such as subsea monitoring, very little literature has addressed this problem specifically.

To bridge the existing gap in the literature, the remainder of this dissertation provides approaches for single UW image super-resolution and deblurring as well as detection and elimination of floating particles in single images and video frames.

In order to restore single UW images, dictionary learning together with the powerful tool of sparse representation, is used. Using a training data set, the relationship between the low-resolution images with their high-resolution counterpart is learned via sparse representation. Employing this information, one can recover the high-quality counterpart of any low-quality input image of the same class. To this end, a short introduction of sparse representation is introduced in the next chapter which is followed by the proposed super-resolution and deblurring approach.

4 Sparse Representation: Definition and Applications

4.1 Introduction

Research in the area of image processing has had a rapid growth in a wide range of applications including inpainting, denoising, super-resolution and compression. A lot of these algorithms use the advantage of sparse representations and achieve state-of-the-art results.

The problem of sparse representation consists of two main parts:

- A generally over-complete basis called a dictionary,
- An algorithm which selects basis vectors (atoms) and the sparse coefficients that are produced in order to approximate an input signal.

It is called sparse since it only uses a small number of atoms to represent a signal.

Definition 4.1 *A dictionary is a matrix whose columns are derived from example signals. It is called over-complete when the number of columns are more than number of rows.*

In order to use sparse representation, a signal of interest has to be compressible.

Definition 4.2 *A signal is termed compressible when the signal vectors can be represented sparsely with an acceptable error.*

Signal compressibility depends on the dictionary used to obtain the representation therefore, the choice of dictionary is an important factor. Various authors have proposed new training algorithms which produce dictionaries adapted to a particular signal class. Dictionary atoms are chosen in such a way that the loyalty and sparsity of representing training data are preserved. This leads to achieving strong results due to the fact that the atoms in a learned dictionary are conveyed from natural signal examples. The advantage of using learned dictionaries has been well acknowledged in [AEB06].

Definition 4.3 *Atom: dictionary columns are called atoms where each can individually be used to represent a signal.*

This chapter contains an introduction to the main concepts of sparse coding and dictionary learning, summarizes the leading approaches for these processes and reviews the applications of sparse representation in the area of image processing.

4 Sparse Representation: Definition and Applications

4.1.1 Problem Formulation

According to the Sparseland model [EA06], a signal can be sparsely represented by a vector which has just a few nonzero components. A simple solution to sparsity is to threshold the representation coefficients subject to a certain set of basis functions. Despite its simplicity, such an approach has shown a poor representation quality. A better alternative is to employ a linear combination of atoms by performing a basis search process over the columns of a dictionary.

To be more precise, consider a given over-complete dictionary $D \in \mathbb{R}^{N \times K}$ where n is the dimension of the signal space and K is the number of dictionary's atoms ($K > N$), and a vectorized signal of interest $\vec{x} \in \mathbb{R}^N$. The signal \vec{x} can be represented over dictionary D as:

$$\vec{x} \approx D\vec{\alpha}, \quad (4.1)$$

where $\vec{\alpha}$ is the sparse coding vector.

Definition 4.4 *Sparse coding vector (sparse coefficient vector) is a vector $\in \mathbb{R}^K$ composed of mostly zero elements and a few nonzeros containing weights corresponding to the used atoms in the dictionary to represent the signal of interest.*

Such a system is known to be ill-posed and has many solutions. Not only a unique solution is required but also it has to have the minimum number of nonzero components (as sparse as possible). It can be described via the two following optimization problems where representation error tolerance ϵ and sparsity T are forced respectively:

$$\underset{\vec{\alpha}}{\operatorname{argmin}} \|\vec{\alpha}\|_0 \quad s.t. \quad \|\vec{x} - D\vec{\alpha}\|_2^2 \leq \epsilon, \quad (4.2)$$

$$\underset{\vec{\alpha}}{\operatorname{argmin}} \|\vec{x} - D\vec{\alpha}\|_2^2 \quad s.t. \quad \|\vec{\alpha}\|_0 < T, \quad (4.3)$$

here $\|\cdot\|_0$ counts the number of non-zero coefficients in a vector where $\|\cdot\|_2$ denotes the Euclidean norm. Basically, sparse approximation is a vector selection problem and requires the availability of a dictionary. In order to solve such a problem, all possible combinations of the atoms should be attempted to find the best combination. However, the complexity of this process is intractable and is known as non-deterministic Polynomial-time (NP)-hard.

In order to find a solution which is more feasible in the sense of complexity, various authors have developed different algorithms. According to the literature, sparse approximation approaches are categorized into greedy algorithms and convex relaxation approaches. In greedy algorithms (matching pursuit family), l_0 norm is minimized in an iterative manner while successively approximating the signal. On the contrary, convex relation approaches replace the l_0 norm with l_1 norm, giving a significant reduction to the computational complexity of the process. Focal Underdetermined System Solver [GR97; ERK00] and Basis Pursuit approach [CDS01] fall in this category.

Given a dictionary D , first, some of atom selection algorithms are introduced. Later, the dictionary learning algorithms are reviewed.

4.1.2 Matching Pursuit Family (MPF)

Greedy sparse approximation algorithms known as the matching pursuit family, provide an approximation solution of 4.2. This is done by minimizing the l_0 norm in an iterative manner. At each iteration i , one atom d_j , $1 < j < K$, is drawn from the dictionary $D = [\vec{d}_1, \dots, \vec{d}_K]$. The selected atom is the one that most correlates to the approximation error. The correlation is measured as the length of the orthogonal projection.

The Matching Pursuit (MP) algorithm proposed by Mallat and Zhang [MZ93] forms the foundation for greedy algorithms based on which other extension and variants such as Orthogonal Matching Pursuit (OMP) [PRK93] and Optimized Orthogonal Matching Pursuit (OOMP) [RL02] were developed.

Matching Pursuit (MP)

The first greedy algorithm known as MP [MZ93] obtains an efficient approximation of Equation (4.2). For this, it first defines a residual vector \vec{r} which is initialized with the signal \vec{x} , ($\vec{r}_0 = \vec{x}$). This vector is in charge of choosing the updated atom-index/coefficient pair. At every iteration i , the residual vector from the previous iteration (\vec{r}_{i-1}), together with a chosen single atom \vec{d}_j from the dictionary D that most correlates to this residual, are used to calculate the sparse coefficient $\vec{\alpha}_i$ as follows:

$$\underset{\vec{d}_j}{\operatorname{argmax}} | \langle \vec{d}_j, \vec{r}_{i-1} \rangle |, \quad (4.4)$$

$$\vec{\alpha}_i = \langle \vec{d}_j, \vec{r}_{i-1} \rangle, \quad (4.5)$$

here $\langle \cdot \rangle$ denotes the inner-product operation defined as $\langle u, v \rangle = u^T v$. Once the best atom and its corresponding coefficients are obtained, the algorithm updates the residual for the next iteration as follows:

$$\vec{r}_i = \vec{x} - \hat{\vec{x}}_{i-1}, \quad (4.6)$$

where

$$\hat{\vec{x}}_{i-1} = \vec{d}_j \vec{\alpha}_i. \quad (4.7)$$

This process repeats until either the norm of the residual falls below the desired approximation error bound or the number of employed atoms reaches the sparsity limit. The approximation at the iteration i is then obtained as:

$$\hat{\vec{x}}_i = \sum_{k=1}^i \alpha_k d_k = [d_1 \ d_2 \ \dots \ d_i] [\alpha_1 \ \alpha_2 \ \dots \ \alpha_i]^T. \quad (4.8)$$

The MP atom/coefficient selection rules have low complexity, and guarantees the convergence since the approximation error decreases in each iteration. However, it suffers from the problem that the residual norm $|\vec{r}_i|$ at the output of iteration i will never be zero even when there are more coefficients $\vec{\alpha}_i$ than the dimension of \vec{x} ($i \geq N$). It can be zero if an atom exist which is collinear with \vec{r}_{i-1} but its probability is zero. Furthermore, it is sub-optimal. This means that, at each iteration the residual is orthogonal only to the immediately selected atom and it may not be orthogonal to the previously selected atoms. This results in slow convergence since some atoms selected at an earlier iteration may get selected again.

Orthogonal Matching Pursuit (OMP)

Orthogonal matching pursuit (OMP) [PRK93] overcomes the above difficulties by assuring the orthogonality of all selected atoms up to the current iteration to the newly derived residual. Thus, once an atom is selected, it will never be selected again.

The same as MP, OMP iterates to find the best atom to represent the residual. For this, at every iteration i , using all the selected atoms up to this iteration it spans a subspace. Employing this subspace it first computes:

$$\underset{\vec{d}_j \in D/D_{i-1}}{\operatorname{argmax}} \quad | \langle \vec{d}_j, \vec{r}_{i-1} \rangle |, \quad (4.9)$$

here r_{i-1} denotes the residual at the $(i-1)$ th iteration, D_{i-1} stands for the set of atoms selected up to the $(i-1)$ th iteration, and $'/'$ defines the set difference operator. Let $D_i \equiv D_{i-1} \cup d_j$, the approximation at iteration i is then computed as:

$$\hat{x}_i = D_i (D_i^T D_i)^{-1} D_i^T \vec{x} \equiv D_i \vec{\alpha}_i. \quad (4.10)$$

where $D_i^\dagger = (D_i^T D_i)^{-1} D_i^T$ is the pseudo-inverse of D_i . The next step is to update the residual for the following iteration:

$$\vec{r}_i = \vec{x} - D_i \vec{\alpha}_i. \quad (4.11)$$

While MP only considers an iteration's input residual to obtain the i th atom/coefficient pair, OMP looks back to all previous iterations and optimizes the coefficients for previously selected atoms. This means that, OMP selects each atom of the dictionary only once and therefore, has a lower degree of freedom in the selection process. Thus, it leads to a better approximation quality compared to MP, thus results in the following property:

$$\hat{\vec{x}}_N = \vec{x}. \quad (4.12)$$

This follows from the fact that D is assumed to be full-rank thus, the Equation (4.9) guarantees that the selected atoms are linearly independent. As before, OMP terminates if the norm of the residual falls below the desired approximation error bound, or if the number of atoms in the approximation reaches the desired limit. Otherwise, it proceeds to the next iteration. The OMP appears to be very simple but, at the same time, very powerful. In this work the OMP approach is used, therefore, it will be discussed in more detail in section 4.2.2.

Optimized Orthogonal Matching Pursuit and other Greedy Algorithms

The Optimized Orthogonal Matching Pursuit (OOMP) [RL02] is an extension of OMP. The same as OMP, OOMP looks back to the previous iterations to modify the coefficients, but it differs from OMP in atom selection. Let $D_{(i-1,j)}$ denotes the matrix containing all the atoms up to the $(i-1)$ th iteration augmented by the j th atom d_j where $d_j \in D/D_{i-1}$, i.e. $D_{(i-1,j)} \equiv [D_{i-1} \quad d_j]$. At the i th iteration, OOMP computes:

$$\underset{\vec{d}_j \in D/D_{i-1}}{\operatorname{argmax}} \quad |D_{(i-1,j)} (D_{(i-1,j)}^T D_{(i-1,j)})^{-1} D_{(i-1,j)}^T \vec{x}|. \quad (4.13)$$

Here, the signal vector is projected onto the subspace spanned by the selected atoms up to the i th iteration and the atom under consideration. The atom with largest projection is then chosen as the optimal solution. Similar to OMP, OOMP selects the linearly independent atoms, the approximation is then obtained with respect to the selected atom and updates the residual as in 4.11. The stopping criterion is the same as the other MPF algorithms, it terminates if the norm of residual is below the desired approximation error bound or the number of atoms in approximation reaches the sparsity limit.

Similar to OMP, OOMP enjoys the exact reconstruction property 4.12 but due to the projection with respect to each candidate atom, it is computationally much more expensive than OMP, and the gain in performance over OMP is not significant.

Along the line of greedy algorithms, various other approaches offer improvement in both representation fidelity and computational complexity such as [DMA97; CBL89; CB00; GMS03]. The aim of all these methods is to provide a more efficient convergence. In signal processing literature, their setting goes by the name of MP and OMP and their several other variants and extensions.

4.1.3 Convex Relaxation Algorithms

The difficulty in solving 4.2 is in l_0 norm minimization. Convex relaxation algorithms get around this problem by replacing the l_0 norm with l_1 norm. This replacement reduces the computational complexity. Moreover, it allows the transformation of this problem into linear programs whose solutions are straight forward. [BDE09].

Basis Pursuit

The Basis Pursuit approach [OF96] is a well-known example of replacing the l_0 norm with the l_1 norm constraint. The resulting algorithm is expressed as the solution to the following problem:

$$\underset{\vec{\alpha}}{\operatorname{argmin}} \|\vec{\alpha}\|_1 \quad s.t. \quad \vec{x} = D\vec{\alpha}, \quad (4.14)$$

the solution generally comprises n linearly independent vectors which are uniquely specified by \vec{x} . A sparse adaptation of this approach can be obtained by minimizing the reconstruction error subject to an inequality constraint on the sparsity; this can be formulated in Lagrangian form as:

$$\underset{\vec{\alpha}}{\operatorname{argmin}} \|\vec{x} - D\vec{\alpha}\|_2 + \lambda \|\vec{\alpha}\|_1. \quad (4.15)$$

Clearly, these problems are different than the original problems in Equations 4.2 and 4.3. Although, solving the l_1 norm minimization results in an approximate solution of l_0 norm minimization problem. It has been shown that if the number of nonzero components in the sparsest solution is less than a certain bound, then the minimum l_0 norm solution is unique and it coincides with the l_1 norm solution [DH01; Don06; EB02]. It is highly dependent on the number of nonzero components and the bound which is related to the dictionary in use.

Focal Under-determined System Solver (FOCUSS)

Another note-worthy algorithm is the focal underdetermined system solver (FOCUSS) which is an approximation algorithm to find the solution for (4.2) by minimizing the l_p norm for $p < 1$ instead of the l_0 norm [GR97; ERK00]. The solution can be obtained by solving following equation:

$$\underset{\vec{\alpha}}{\operatorname{argmin}} \|\vec{\alpha}\|_p^p \quad s.t. \quad \vec{x} = D\vec{\alpha}. \quad (4.16)$$

This problem is non-convex for $p < 1$, but a locally optimal solution can be achieved using Lagrange multipliers.

There are other less well-known algorithms for sparse coding in the literature such as [OF97; LO99] which are based on the *maximum a posteriori* (MAP) estimation of the solution vector $\vec{\alpha}$. Assuming some prior distribution for $\vec{\alpha}$, it maximizes the *a posteriori* probability $\mathbb{P}(\vec{\alpha}|\vec{x}, D)$. It tries to solve the sparse model when the observation is accompanied by random noise.

4.1.4 Choice of Dictionary

The introduced sparse approximation problems above require the availability of a dictionary D . Knowing the alternatives to solve such a problem, one needs to choose the proper dictionary since it plays an important role in the success of the sparse approximation problem. One common assumption when employing sparse representations is that the signal vectors in question are compressible. This means that the signal vectors can be well-represented using an over-complete dictionary while it uses only a small number of dictionary atoms. Thus, signal compressibility depends on the dictionary which is used to obtain the representation.

There are two alternatives for the choice of dictionary:

- Fixed dictionaries: these so-called off-the-shelf dictionaries are the standard algebraic basis functions such as Gabor, wavelet, curvelet and Fourier basis functions. Despite their simple usage, they do not fit with a large set of signals [SE10].
- Learned dictionaries: these dictionaries are obtained over a training set and have shown more adaptability to the structure of the signal, and have a better capability to signal-fitting [SE10].

Definition 4.5 *DL: the process of learning a dictionary over a set of available training signals such that it is well-adapted to its purpose is termed dictionary learning (DL).*

A learned dictionary should possess two characteristics: first, its loyalty at representing the data which it is trained on. Second, the sparsity of such a representation.

Consider a set of training signals $X = [\vec{x}_1, \vec{x}_2, \dots, \vec{x}_M] \in \mathbb{R}^{N \times M}$, the DL aims at learning a dictionary $D = [\vec{d}_1, \vec{d}_2, \dots, \vec{d}_K] \in \mathbb{R}^{N \times K}$ which represents the training samples with coefficients

$A = [\vec{\alpha}_1, \vec{\alpha}_2, \dots, \vec{\alpha}_M] \in \mathbb{R}^{K \times M}$. The following minimization problem formulates the DL problem through the cost function $f(\cdot)$ as:

$$\underset{D, A}{\operatorname{argmin}} f(D, A) = \underset{D, A}{\operatorname{argmin}} \|X - DA\|_F^2. \quad (4.17)$$

The DL problem can be conducted as an iterative process over two stages of learning the dictionary D and calculating the sparse coefficients A until it achieves the optimal solution:

- Given the dictionary D , the sparse coefficients A are calculated
- Updating the dictionary D using calculated sparse coefficients

DL algorithms differ in the method of finding the sparse coefficients and procedure of updating the dictionary.

Olshausen and Field proposed one of the earliest dictionary training algorithms [OF96]. They learned an optimal dictionary using maximum-likelihood estimation. While estimating the optimal dictionary, a Gaussian or Laplace prior is considered on the sparse representation coefficients. They employed the steepest descent method to update both dictionary and the sparse coefficients.

Following this approach, several other DL algorithms have been proposed such as [Eng99; CR02; Kre+03; AEB06; ESH07; Mai+08; YBD09; Mai+09a] and [SE10]. Here is a brief revision of some benchmark DL approaches.

The Method of Optimal Directions (MOD)

The Method of Optimal Direction (MOD) [EAH99] proposed by Engan et al. instead uses a simple dictionary update procedure while using OMP [PRK93] or FOCUSS [GR97] algorithm in the sparse coding stage. The best dictionary solution to the quadratic function

$$\|X - DA\|_F^2, \quad (4.18)$$

can be obtained by forcing its derivative, respect to D , to zero when the sparse coefficients A are assumed to be fixed:

$$D = XA^T(AA^T)^{-1}, \quad (4.19)$$

thus, solving the dictionary update stage in one step. Similar to other DL algorithms, the MOD approach alternates between the dictionary update stage and sparse approximation stage. Despite its simplicity at updating the dictionary, the MOD suffers from high complexity of the matrix inverse, particularly for large dictionaries.

Unions of Orthogonal Matrices

One alternative to reduce the complexity and simplify the training task has been introduced in [SHG08; Les+05] where dictionaries are considered to be the unions of orthonormal bases $D_k \in \mathbb{R}^{N \times N}$, $k = 1, \dots, L$:

$$D = [D_1 \mid \dots \mid D_L]. \quad (4.20)$$

4 Sparse Representation: Definition and Applications

The coefficients of sparse representation A are decomposed to the same number as orthonormal bases and every of them correspond to a different bases. Such a configuration simplifies the sparse coding stage. The proposed algorithms update every orthonormal matrices sequentially using singular value decomposition.

Sezar et al. [SHG08] proposed such an algorithm where it assumes that the sparse decomposition process is constrained so that the representation uses a single basis D_k from D which is chosen based on a sparsity criterion. Letting $\vec{\alpha}_k$ denote the representation that uses orthonormal bases D_k , the l_0 constrained sparse representation is expressed as follows:

$$\underset{\vec{\alpha}_k}{\operatorname{argmin}} \|\vec{x} - D_k \vec{\alpha}_k\|_2^2 + \lambda \|\vec{\alpha}_k\|_0. \quad (4.21)$$

In every iteration D_k is updated and then used to re-classify the training set and obtain new $\vec{\alpha}_k$ and \vec{x} . For more details, the reader is referred to the cited articles.

One alternative approach to the constrained sparse representation setup presented above, [Les+05], considers the structure in 4.20 but does not constrain the decomposition scheme to use a single D_k . It instead uses an algorithm such as a matching pursuit applied over the entire dictionary D . The sparse representation of $X = \operatorname{cols}(\{\vec{x}\})$ is written as:

$$DA = [D_1 \mid \dots \mid D_L] \begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ A_L \end{bmatrix} = \sum_k D_k A_k. \quad (4.22)$$

The dictionary update then proceeds by modifying one D_l at a time while keeping all other bases $D_k, k \neq l$, fixed. Let

$$E_l = X - \sum_{k \neq l} D_k A_k \quad (4.23)$$

be the error matrix inside 4.18 with the term $D_l A_l$ removed. Each updated D_l is chosen to minimize

$$\|E_l - D_l A_l\|_F^2. \quad (4.24)$$

KSVD Dictionary Learning

Aharon et al. [AEB06] proposed a method termed K-SVD which differs from the approaches discussed above where they freeze A while updating the dictionary D and vice-versa. The K-SVD algorithm updates the dictionary sequentially and lets the relevant coefficient change as well. The solution follows from the singular value decomposition (SVD) of E_l in Equation 4.23.

The joint dictionary learning and sparse representation of a signal can be defined by the following optimization problem:

$$\underset{D, A}{\operatorname{argmin}} \|X - DA\|_F^2 \quad s.t. \quad \forall i, \|\vec{\alpha}_i\|_0 < T, \quad (4.25)$$

where $\|\cdot\|_F^2$ denotes the Frobenius norm.

Definition 4.6 *Frobenius norm of a matrix is defined as the square root of the sum of the absolute squares of its elements.*

The K-SVD algorithm first uses an initial estimation for dictionary and, by employing a pursuit algorithm, calculates the best atoms from the current dictionary to represent data X . Then, with the representation coefficients calculated, it updates both the dictionary and the representation coefficient as well. In every iteration, just one atom is replaced in the updated dictionary and it is selected such that reduces the error or, in the worst situation, it remains the same as the previous iteration. This approach will be discussed in more detail in section 4.2.1.

Online Dictionary (ONLD)

The DL algorithms discussed above cannot handle a very large training set efficiently since they access the entire training set at each iteration. To overcome this challenge, a new scheme, called online dictionary learning algorithm (ODL), has been introduced [Mai+09a; Mai+10]. Such approaches are generally based on Stochastic Gradient Descent (SGD) method and, instead of training over the entire set as a batch, they use a single or a small number of training examples at a time to process. This leads to a lower memory requirement and faster convergence rates [Mai+10].

One interesting example of such an approach is the Image Signature Dictionary (ISD) proposed by Aharon et al. [AE08]. Based on this approach, in order to have a more compact dictionary, some of the dictionary redundancies are sacrificed. Thus such a method enjoys having a reduced training complexity and consequently becomes an interesting approach for online tasks.

Based on this approach, dictionary is represented as a small image of N pixels where the size of dictionary assumed to be $N \times K$. Every one of N pixels of the ISD is surrounded by dictionary atoms $d_k \in \mathbb{R}^N$ as a block. The dictionary update is a two-step process where either the sparse representations $\tilde{\alpha}$ or the dictionary D are getting updated when the other one is fixed.

A variation of this approach is proposed by Mairal et al. [Mai+10; Mai+09a]. The same as K-SVD approach [AEB06], they employed a block-coordinate descent in the dictionary update stage. Instead of employing a first order linear model to handle the SGD dictionary update, the method updates one atom at a time while holding the remaining atoms and all sparse coefficients fixed. The minimization of the following objective function continues until convergence:

$$f_T(D) = \frac{1}{T} \sum_{t=1}^T \|\vec{x}_t - D\tilde{\alpha}_t\|_2^2 + \lambda \|\tilde{\alpha}_t\|_1, \quad (4.26)$$

at the current time T a single sample \vec{x}_T is drawn from the training set (or training stream), decomposed using D_T to produce $\tilde{\alpha}_T$ and then, both \vec{x}_T and $\tilde{\alpha}_T$ are used to update the constant coefficients of the quadratic expression 4.26.

4.1.5 Application of Sparse Representation using Dictionary Learning

There is a wide range of sparse representation applications and in this section we briefly describe some of them.

Inpainting

Image inpainting is a useful application in several scenarios of image processing. It is used to fill in pixels which are missing in an image. In the context of data transmission, it can be used to provide an alternative to channel codes [ZL06; RG04] commonly used to model packet-based network transmissions. Examples of inpainting in image manipulation include the removal of superposed text, road-signs or publicity logos [Ela+05].

Considering an image patch $x = [\vec{x}_a^T \ \vec{x}_m^T]^T$ which is made of two sub-vectors, sub-vector \vec{x}_a is defined for the available pixels in the patch and \vec{x}_m stands for the missing pixels where image inpainting is used to estimate. Guleryuz [Gul06] proposed a method to estimate the missing sub-vector \vec{x}_m . They approximate the missing data employing a concatenation of orthonormal bases that render \vec{x} compressible. Compressibility of a signal \vec{x} means that given D , there exists some sparse vector $\vec{\alpha}$ which satisfies $\vec{x} = D\vec{\alpha}$.

Without loss of generality, assuming $\vec{\alpha} = [\Gamma^T \ 0^T]^T$, and using the orthonormality of D , $\vec{\alpha}$ can be written as:

$$\vec{\alpha} = \begin{bmatrix} \Gamma \\ 0 \end{bmatrix} = D^T \vec{x} = \begin{bmatrix} D_{a,n} & D_{m,n} \\ D_{a,z} & D_{m,z} \end{bmatrix} \begin{bmatrix} \vec{x}_a \\ \vec{x}_m \end{bmatrix}, \quad (4.27)$$

where subscripts n and z stand for non-zero and zero components of $\vec{\alpha}$ respectively. The sparsity constraint

$$0 = D_{a,z}\vec{x}_a + D_{m,z}\vec{x}_m \quad (4.28)$$

can be used to extract one possible estimation of missing data as follows:

$$\hat{\vec{x}}_m = -(D_{m,z}) + D_{a,z}\vec{x}_a. \quad (4.29)$$

The problem with this approach lies in the estimation of the support of $\vec{\alpha}$ which without it, partitioning D^T as in 4.27. Instead one can build an estimation of $\hat{\vec{x}}_m$ iteratively by subsequently enforcing:

- an estimation of the sparsity constraint.
- the available data constraint.

Let the diagonal matrices Δ_n , Δ_a and Δ_m be diagonalized with 1/0-value to define the non-zero entries of $\vec{\alpha}$, the available and missing entries of \vec{x} respectively. Thus, the i -th estimation of \vec{x} can be expressed as:

$$\hat{\vec{x}}^i = \begin{bmatrix} \vec{x}_a \\ \hat{\vec{x}}_m^i \end{bmatrix} = \Delta_a \hat{\vec{x}}^{i-1} + \Delta_m (D \Delta_n D^T) \hat{\vec{x}}^{i-1}, \quad (4.30)$$

Guleryuz showed that it converges to 4.29 where the iterative nature of the 4.30 permits an adaptive selection of the support of $\vec{\alpha}$ wherein the support is re-constrained after every few iterations.

Denoising

Sparse representations have also been used to denoise the images and videos [EA06; EA06; MES08; PE09]. The denoising problem is formulated as a MAP estimation problem when a sparse prior is considered on the data. By obtaining the sparse estimation of image blocks which are overlapped, the solution of MAP approximation is defined; and then the denoised data is specified by averaging over all blocks.

Consider a noisy image of size $N \times M$ with the representation X_n . All unique overlapping $b \times b$ blocks are extracted and then reshaped to vectors and stacked together to form a large vector $Y \in \mathbb{R}^{(N-b+1) \cdot (M-b+1)}$. Accordingly write the 1/0-valued matrix Ω with i -th row specifying all positions in Y where pixel N_i occurs. \hat{Y} can be build by replacing each block y in Y by its sparse approximation $\hat{y} = D\tilde{\alpha}$ over an over-complete dictionary D so that the error norm $|y - \hat{y}|$ is similar to the noise variance. Thus the denoised image estimate \hat{X}_n proposed in [EA06; EA06] is built as:

$$\hat{N} = \frac{\Omega\hat{Y} + \lambda X_n}{\Omega\Omega^T + \lambda}, \quad (4.31)$$

here the denominator of this expression is a diagonal matrix with i -th diagonal entry given by $n_i + \lambda$ where n_i is the number of unique and overlapping blocks in the image that contains the i -th pixel. Hence the above expression just specifies an averaging of all sparse approximations of any given pixel with a relaxation using the noisy image X_n .

Texture Separation and Classification

There are a series of works using the application of sparse representations to texture separation [SED05; SED04; PFS10]. According to this application, every image block x is assumed to consist of a mixture of overlapping component layers \vec{u}_k :

$$\vec{x} = \sum_k \tilde{\alpha}_k \vec{u}_k, \quad (4.32)$$

this model coincides with the case of mixtures between multiple images or to natural images consisting of a cartoon layer with superposed texture. Sparse representation can be adapted to this problem if assuming the available dictionaries D_k which can render \vec{u}_k compressible. In order to use this tool, two issues arise:

- Obtaining the dictionaries D_k .
- Using the dictionaries to separate the various layers of the image.

According to this approach, two kinds of dictionaries are used for edge/contours textures layers (named cartoon layers) and one for more complex texture layers (oscillatory texture layers). The cartoon layers are modeled using a combination of off-the-shelf dictionaries (fixed for all the image patches) and also the learned, adaptive dictionaries are employed in order to sample the more complex texture layers. Then the problem is solved in three stages:

- Calculating the sparse coefficient $\tilde{\alpha}$ by using linear programming methods.

4 Sparse Representation: Definition and Applications

- Employing conjugate gradient descent in order to solve texture layers.
- Solving the adaptive dictionary using the gradient descent method.

Image Compression

Recently, in the case of image compression, using learned over-complete dictionaries which are adapted to a signal class, lead to successful result. The advantage of using a learned dictionary is achieving a greater compressibility of the considered signal class. An example of such an approach is the work which is introduced by Bryt and Elad [BE08] based on the learned K-SVD dictionary [AEB06]. In their approach, they used some pre-specified face templates. Each of the face templates which are not overlapped with the others, are used in order to specify a class of signals. These classes are then represented by employing the corresponding K-SVD dictionary. This results in a wide PSNR improvement over the state-of-the-art JPEG2000 [SCE01] algorithm but, at the same time, it suffers from the expense of large storage for the dictionaries.

Image Super-Resolution

The problem of super-resolution is an important active area of research due to wide demand for high-resolution images in many applications. Obtaining a high-resolution (HR) image from a single or multiple low-resolution (LR) images, that have lost their higher frequency information during acquisition, transmission or storage, is known as super-resolution. Conventional SR methods require using several LR images to reconstruct the HR one. In situations where the number of available input images is small, these algorithms are not considered to be practical. In such cases Single Image Super-Resolution (SISR) draw more attention. The most recent and successful approaches to this purpose use sparse representation to enhance the quality of an image.

Such a problem is coupled feature space meaning that it deals with the high-resolution (HR) and low-resolution (LR) space at the same time [Yan+08]. Intuitively, a single sparse coding model is used for learning dictionaries in a coupled feature space [Yan+08]. However, these dictionaries could not capture the complex, spatial-variant and the non-linear relationship between the two feature spaces.

Many approaches have been proposed to capture the concurrent prior between the LR and HR patches. R. Zeyde et.al [ZEP10] have proposed a two-step learning approach. First, a LR dictionary is learned using K-SVD [AEB06] and then the HR counterpart is obtained via the least square error calculations. Although the dictionaries are learned individually, the same sparse coefficients are used for both feature spaces. [Yan+12a; Yan+10] use the same sparse coefficients in both spaces and move one of the optimization problem to the regularization term of the other one. Thus, it avoids solving two optimization problems.

Wang et al. [Wan+12] proposed a semi-coupled dictionary learning model which uses a mapping matrix to capture the relationship of the sparse representation between spaces. This leads to a better minimization error in both spaces rather than those which are learned in concatenated spaces, but the corresponding relationship of dictionaries in two feature spaces are not captured

during the learning process.

Having the corresponding LR and HR dictionaries, super-resolution employs one of the atom selection algorithms and calculates the sparse coefficients of the LR input image. Once the sparse coefficients are available, together with the HR dictionary the HR output image is conducted in a patch-based manner [Yan+08; Wan+12; Yan+10; Sun+03; CYX04]. The SR algorithm in this thesis is in line with Zeyde's method which is explained in the next section.

4.2 Single Image Super-Resolution (SISR)

Obtaining a high-resolution (HR) image X from the single low-resolution (LR) image Y is known as single image super-resolution (SISR). The LR image is a version of the HR image which has lost its higher frequency information during acquisition, transmission or storage. In order to solve such an ill-posed inverse problem via sparse representation, two constraints are assumed:

- Reconstruction constraint: based on image observation model, reconstructed HR image X should be in agreement with the LR image;
- Sparsity prior: HR image can be sparsely represented over an over-complete dictionary and it can be recovered from the LR version.

To be more precise, consider LR image Y which is downsampled and blurred version of HR image X . Further, assume that there is a HR over-complete dictionary $D_h \in \mathbb{R}^{N \times K}$ of K bases which is a large matrix learned using HR image patches. Then the vectorized patches of image X , $\vec{x} \in \mathbb{R}^N$ can be sparsely represented over dictionary D_h . So the high resolution patch \vec{x} can be represented as $\vec{x} = D_h \vec{\alpha}_0$ where $\vec{\alpha}_0 \in \mathbb{R}^K$ is a vector with very few nonzero elements ($\ll K$). The relationship between an HR image patch \vec{x} and its LR counterpart \vec{y} can be expressed as:

$$\vec{y} = SB\vec{x} = L\vec{x} = LD_h\vec{\alpha}_0, \quad (4.33)$$

note that S is representing a down sampling operator, B representing a blurring filter and L denotes their combined effect. Substituting the representation for the HR image patch \vec{x} into 4.33 and noting that $D_l = LD_h$, one gets:

$$\vec{y} = LD_h\vec{\alpha}_0 = D_l\vec{\alpha}_0. \quad (4.34)$$

The above equation implies that the LR image patch \vec{y} will also have the same sparse representation coefficient $\vec{\alpha}_0$. Now, given the LR image patches, one can obtain the representation coefficients using a vector selection algorithm such as OMP (Orthogonal Matching Pursuit). After obtaining sparse coefficients, one can reconstruct the high resolution and deblurred image patch \vec{x} as follows:

$$\hat{\vec{x}} = D_h \hat{\vec{\alpha}}_0. \quad (4.35)$$

The sparse representation problem (vector selection) has the formulation as an optimization problem which results in finding the sparse coefficient $\vec{\alpha}$ using dictionary D_l . To obtain the sparse representation coefficient for the LR image patch \vec{y} , one solves the following optimization problem:

$$\min_{\vec{\alpha}_0} \|\vec{y} - D_l \vec{\alpha}_0\| \quad \text{subject to} \quad \|\vec{\alpha}_0\| < T, \quad (4.36)$$

4 Sparse Representation: Definition and Applications

where T is a threshold which is used to control the sparseness of the representation. The l_0 norm is used to identify the number of nonzero elements of the vector $\vec{\alpha}_0$.

In order to represent the signal of interest, a suitable dictionary and a sparse linear combination of the dictionary atoms are needed. The sparse representation problem subject to find the most proper selection of those linear combination vectors from an over-complete dictionary D_I . To find such a representation, different pursuit algorithms can be used such as OMP and the over-complete dictionary can be formed using K-SVD.

4.2.1 K-SVD Approach

As was mentioned previously, an over-complete dictionary together with the sparse coefficients are needed to represent a signal. The joint dictionary learning and sparse representation of a signal can be defined by the following optimization problem:

$$\min_{D_I, A} \|Y - D_I A\|_F^2 \quad \text{subject to} \quad \forall \|\vec{\alpha}_i\| < T. \quad (4.37)$$

Consider a set of over-complete basis vectors Y , and an initial dictionary which is formed by choosing its elements from the training set randomly, D_I . The K-SVD algorithm involves two basic steps, which together constitute the algorithm iteration:

- The signal in X are sparse coded given the current dictionary estimate, producing the sparse representation matrix A ,
- The dictionary atoms are updated given the current sparse representation.

Algorithm 1 summarizes the steps. The sparse coding is commonly done using OMP via following optimization problem for each and every input signal \vec{y}_i :

$$\min_{\vec{\alpha}_i} \|\vec{y}_i - D_I \vec{\alpha}_i\|_2^2 \quad \text{subject to} \quad \|\vec{\alpha}_i\|_0 \quad i = 1, 2, \dots, N. \quad (4.38)$$

Since K-SVD algorithm attempts to update the dictionary by replacing one atom at a time to reduce the error in representation, thus in every iteration, the dictionary and the effective sparse coefficient vectors are considered to be fixed and just one atom in the dictionary is questioned to be replaced and the corresponding sparse coefficient is then calculated. The main innovation in the algorithm is the atom update step, which is performed while preserving the constraint in 4.37. For this, the update step uses only the signals in X whose sparse representations use the current atom. Letting I denote the indices of the signals in X which use the j -th atom, the update is obtained by optimizing the target function

$$\|X_I - D A_I\|_F^2 \quad (4.39)$$

over both the atom and its associated coefficient row in A_I . The resulting problem is a simple rank-1 approximation task given by

$$\{\vec{d}, \vec{g}\} := \underset{\vec{d}, \vec{g}}{\text{Argmin}} \|E - \vec{d} \vec{g}^T\|_F^2 \quad \text{subject to} \quad \|\vec{d}\|_2 = 1, \quad (4.40)$$

Algorithm 1 K-SVD Dictionary Learning Algorithm**Input:** Training set $X \in \mathbb{R}^{N \times M}$, initial dictionary D_0 , target sparsity T and number of iteration k .**Output:** Dictionary D and sparse matrix A such that $X = DA$ Initialization: $D := D_0$ **for** $n = 1 \dots k$ **do** $\forall i : A_i := \text{Argmin}_{\vec{\alpha}} \|\vec{x}_i - D\vec{\alpha}\|_2^2 \quad \text{Subject To} \quad \|\vec{\alpha}\|_0 \leq T$ **for** $j = 1 \dots L$ **do** $D_j := \vec{0}$ $I := \text{indices of the signals in } X \text{ whose representation use } \vec{d}_j$ $E := X_I - DA_I$ $\{\vec{d}, \vec{g}\} := \text{Argmin}_{\vec{d}, \vec{g}} \|E - \vec{d}\vec{g}^T\|_F^2 \quad \text{Subject To} \quad \|\vec{d}\|_2 = 1$ $D_j := \vec{d}$ $A_{j,I} := \vec{g}^T$ **end for****end for**

where $E = X_I - \sum_{i \neq j} \vec{d}_i A_{i,I}$ is the error matrix without the j -th atom, \vec{d} is the updated atom, and \vec{g}^T is the new coefficients row in A_I . The problem is then solved directly using singular value decomposition (SVD), the maximum value modifies the updated atom \vec{d} as the normalized version of the first column of U ($\frac{u_1}{\|u_1\|_2}$) and \vec{g}^T as the first column of V multiplied by $\Delta(1, 1)$ and $(s_1 v_1)$.

4.2.2 Orthogonal Matching Pursuit (OMP)**Algorithm 2** Orthogonal Matching Pursuit**Input:** Dictionary D , signal \vec{x} , target sparsity T or target error ϵ **Output:** Sparse representation $\vec{\alpha}$ such that $\vec{x} \approx D\vec{\alpha}$ Initialization: $I := ()$, $r := \vec{x}$, $\vec{\alpha} := \vec{0}$ **while** (stopping criterion not met) **do** $\hat{k} := \text{Argmax}_k |d_k^T r|$ $I := (I, \hat{k})$ $\vec{\alpha}_I := (D_I)^+ \vec{x}$ $r := \vec{x} - D_I \vec{\alpha}_I$ **end while**

Obtaining an exact sparse representation of a signal is not easily achievable. As a result, many researchers have sought to find the best approximate solution. Among all the methods Orthogonal Matching Pursuit has been the main choice due to its simplicity and fast running time.

4 Sparse Representation: Definition and Applications

Given a dictionary, the greedy OMP algorithm aims to approximate the solution of one of the problems in Equation 4.2 or 4.3. It is an iterative algorithm which updates the basis vector in every iteration by selecting the atom with the largest absolute projection on the error vector. This results in selection of atoms which contain maximum information and consequently reduce the error in the reconstruction. Once the atom is selected, the signal is orthogonally projected to the span of the selected atoms, the residual is computed and the process repeats (see Algorithm 2).

5 Underwater Single Image Super-Resolution and Deblurring via Sparse Representation

5.1 Introduction

Super-resolution algorithms attempt to increase the resolution of input images when the salient image features such as edges, corners, and texture are reconstructed properly. A very common approach for super-resolution is dealing with image features in a patch-based pattern rather than the whole image at once. This helps the reconstruction process, however, it is still challenging. An important discriminating property for image patches is their level of blurriness/sharpness. This property is especially very advantageous when restoring images degraded by spatial varying blur (i.e. UW images).

The popular and well-known DL-based approaches work reasonably well with spatially invariant and mild defocus blur. However, when applying it on an image which is defected by spatially variant blur, some artifacts appear. For example, unwanted over-sharpening effects such as ringing or a halo at the regions with lower blur ratio or in-focus. Figure 5.1 visually illustrate the artifacts. This is due to incapability of these methods to learn a single dictionary which can properly handle varying blur models.

In this case, multiple dictionaries can be considered for different blur levels in an image. Thereby, one can avoid over-sharpening or under-sharpening of the regions with lower levels or higher blurriness respectively.

Furthermore, it is well-known that the success of sparse representation came as a result of employing redundant learned dictionaries. The representation power of learned dictionaries depends on its redundancy. Generally speaking, more redundancy means more possible atoms to be used for a signal approximation. However, redundancy cannot be arbitrarily increased due to two concerns. Firstly, increasing the dictionary redundancy increases the computational complexity and secondly, it associates instabilities and degradation in the sparse approximation process [BDE09]. Therefore, multiple class dictionaries are considered instead of a highly redundant one [Don+11; Fen+11; YSM10; Yan+12b]. This allows for designing compact class dictionaries and leads to high representation quality at reduced redundancy levels. Aside from the computational complexity and stability concern, this setting allows for reducing the degree of freedom for sparse approximation as compact dictionaries are used.

Thus, the work conducted in this chapter comes along the line of a multiple dictionary setting. The major objective is to devise a SISR algorithm using sparse representation together with DL over clusters to restore UW images suffering from spatially variant blur. The proposed algorithm

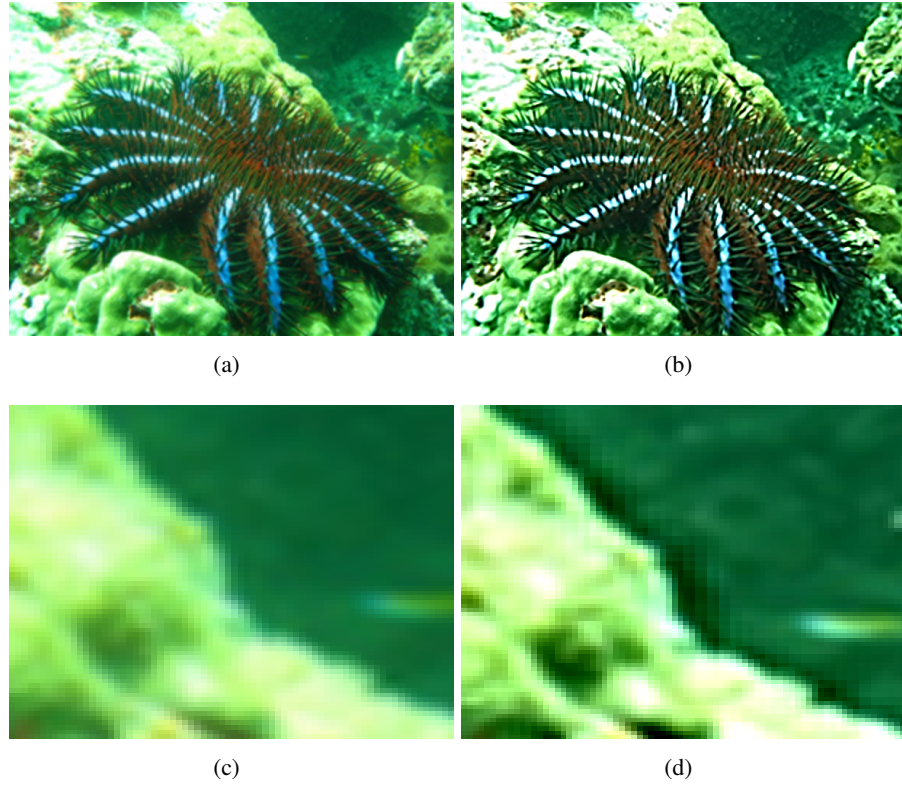


Figure 5.1: An example of artifacts caused using a single dictionary pair. Blurry underwater image (a), reconstructed output estimate (b). Image (b) shows that image is sharpened significantly but halo artifacts also appear in the area with a lower level of blurriness, (c) and (d) are the zoom-in illustration of (a) and (b).

classifies image patches into several clusters based on their level of blurriness and treats them accordingly using a proper dictionary pair. For this, a classifier that can effectively separate image patches is proposed.

In the training stage, using the proposed classifier, the training image patches are grouped into a number of clusters. In essence, the image patches belonging to each cluster inherit the same level of blurriness. This classification serves for designing dictionary pairs that are well suited to represent image features with various blurriness levels. Once the clusters are available, using multiple dictionary learning strategy, a set of structured coupled low-resolution (LR) and high-resolution (HR) dictionary pairs is designed (a pair of dictionaries for each cluster).

In the reconstruction stage, having the dictionary pairs, an input LR image patch is evaluated using the proposed blurriness metric and is categorized into the suitable cluster. The corresponding dictionary pair is then used to restore its HR counterpart.

In addition, an extra representation of blur (in this manuscript UW blur) is applied on the training images to learn dictionaries especially suitable for UW image reconstruction. This helps the algorithm to not only restore the salient properties of an image by super-resolution but also sharpen them (deblurring).

The rest of this chapter is organized as follows: first, a short introduction to sources of blur UW is provided and the degradation model for the blur caused by forward scatter is detailed. Second, the proposed classifier named blur measure (BM) is explained in detail. Third, the proposed SISR and deblurring algorithm in two main subsections, namely the training phase and the reconstruction phase, is described. The experimental results and discussion follows as the next section and finally, the conclusion is provided.

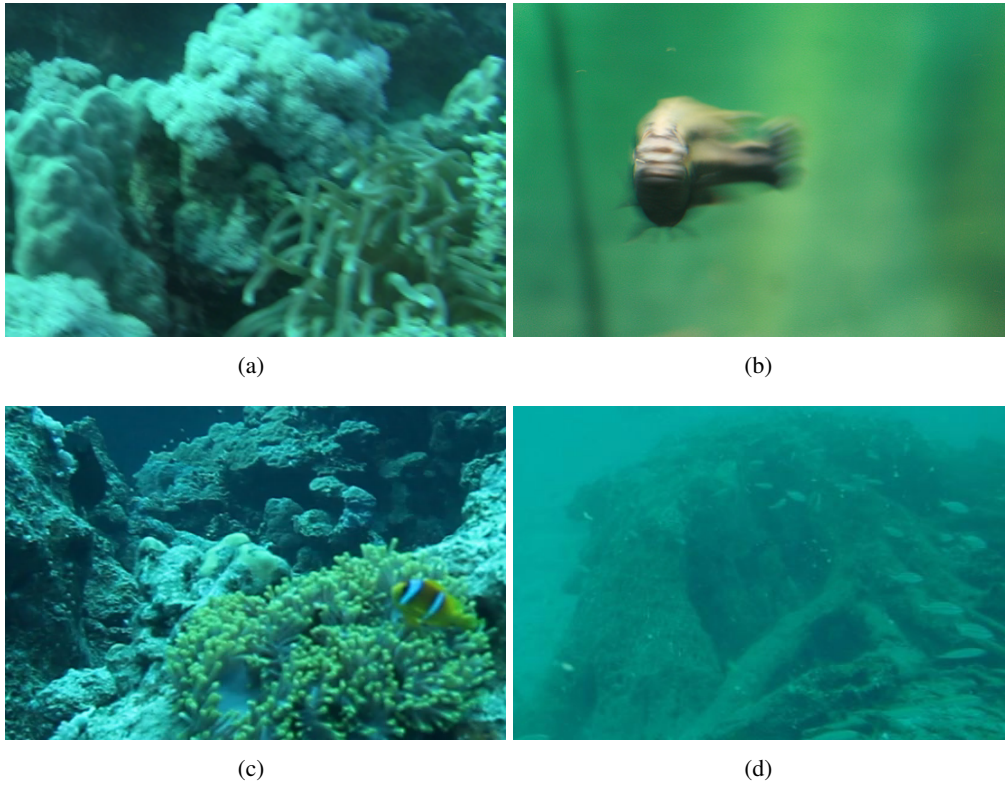


Figure 5.2: Examples of different types of blur underwater: (a) uniform motion blur caused by camera shaking, (b) non-uniform motion blur caused by object movement, (c) defocus blur, (d) spatially varying blur caused by medium.

5.2 Underwater Blur and Its Degradation Model

According to its sources, blur in an UW image can be generally categorized into three groups: motion blur, lens blur, blur due to medium (e.g. scattering) Figure 5.2 illustrates some examples.

Either camera or object movement during the exposure period would lead to motion blur. This phenomenon is common UW where a diver cannot hold the camera sufficiently steady, and thus it is easy to generate camera shaking blur. Fast exposure could reduce the blur amount to some degree. Ideally, if the scene is static with uniform depth and if the camera motion is 2-D translational, then the motion blur can be viewed as spatially invariant (Figure 5.2 (a)) which could be estimated and removed through a blind deconvolution procedure [Fer+06; SJA08]. An alternative is to use robots such as ROV and AUV together with a fast exposure camera.

Incorrect lens setting or limited depth of field would produce defocus blur (example in Figure 5.2 (c)), which is an important type of lens blur. Besides, even if the scene is perfectly in focus and no matter how well the lens is corrected, in most optical imaging systems there is always a fundamental resolution limit due to diffraction, which is called diffraction-limited blur [Bor+99]. Generally speaking, diffraction-limited blur can be approximately viewed as spatially invariant, while the spatial variance of defocus blur depends on both the depth of field of the lens and depth of the scene [FS05; FS03].

Water itself can degrade the performance of related imaging systems [McG80; Jaf90]. Generally speaking, water and its density cause scattering which produces two major distortions: haze and blur, where both are temporally and spatially changing (Figure 5.2 (c)). Removing such effects is very important for many applications and meanwhile it is quite challenging [WL13; TO06; SNN01].

Among aforementioned sources, the most common and disturbing one is the blur due to medium. Unlike in air images, UW images suffer from serious degradation caused by light scattering and absorption. Therefore, this chapter deals with this kind of blur. Following, this source of blur and its degradation model are explained in more detail.

As was mentioned in Chapter 3, an UW image formation consists of three components namely direct, forward scatter and backscattering components. The image corresponding to the direct component is a clear and high-quality image. Forward scatter component is responsible for the small angle scattering effect which leads to a blur image of the scene. Finally backscattering is the one which carries no information from the scene but just reflected light from water molecules and floating particles in the water to the camera and leads to a hazy appearance of the scene. Therefore, blur UW is mainly due to forward scatter. According to the Jaffe-McGlamery image formation, the degradation model for this component can be calculated from the direct component via the convolution relationship:

$$E_{fs}(x, y) = E_d(x, y, 0) * g(x, y | R_c, G, c, B), \quad (5.1)$$

where $*$ represents the convolution operation and

$$g(x, y | R_c, G, c, B) = (e^{-GR_s} - e^{-cR_s}) \mathcal{F}^{-1}(e^{-BR_s f}), \quad (5.2)$$

here R_c is the camera-object distance, the constant c is attenuation coefficient of medium (depending on water type), G is an empirical constant ($|G| \leq c$), B is empirical damping factor and ω is

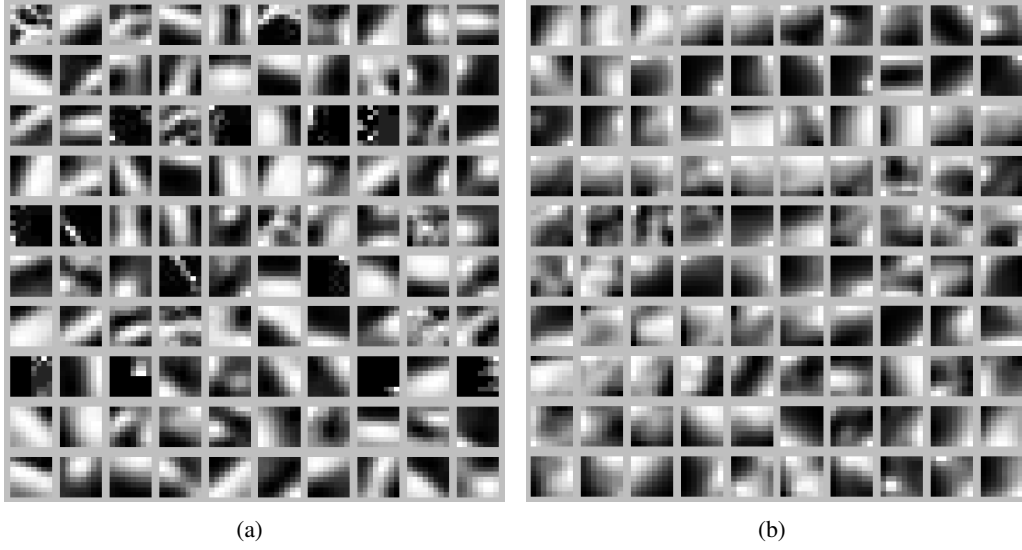


Figure 5.3: Visualizing two dictionaries learned using: (a) clear and sharp underwater image patches, (b) blurred underwater image patches.

spatial frequency in *cycles/rd*. The operator F^{-1} stands for an inverse Fourier transform. Two simplified versions of this model, proposed by Trucco and Olmos-Antillon [TO06] and Later Wu and Li [WL13], are represented as follows respectively:

$$G(\omega, R_c, c, K) \approx Ke^{-cR_c\omega}, \quad (5.3)$$

where $[e^{-GR} - e^{-cR}]$ is approximated with a value between 0.2 and 0.9 termed K , and attenuation coefficient c is considered to be a well approximation of B in forward scatter model and replaced it.

$$G(\omega, k, b) = 1 + k \frac{1 - e^{-b\omega}}{\omega}, \quad (5.4)$$

where $k = k'/B$ and $b = BR_c$.

According to the aforementioned models, it can be seen that this kind of blur which in this thesis (for simplicity) is termed UW blur, is spatially variant depending on the distance R_c . The further the object is from the camera, the more blur appears.

5.3 The Blur Measure (BM)

In order to learn dictionaries with spatially varying blur, one needs a criterion to cluster the training data. Based on [SXJ15] when decomposing local image patches into dictionary atoms, clear and blur dictionaries show quantitatively and visually different results. The diverged effect

manifests that dictionary atoms can precisely characterize structure in blur images, thus amplifying the inherent difference between blur and clear regions.

Figure 5.3 shows two dictionaries learned using (a) clear images and (b) images blurred using the degradation model in Equation 5.4 following the procedure of K-SVD algorithm [AEB06]. The contrast between dictionaries shows how blur, even slightly, influences the fundamental atoms in image decomposition. This observation was the motivation to introduce a blur metric based on dictionary inherent discrepancy between the two types of dictionaries and their ability to represent the data with different level of blurriness.

To this end, first a dictionary D is learned over hundred of thousands of patches cropped from training images which are blurred using the degradation model in Equation 5.4. The dictionary contains 144 atoms of size 36 (the reason for such a choice is explained based on a study in experimental result section). Using this dictionary, a test image data set containing both blur and sharp image patches are decomposed to identify blur:

$$\|\vec{z} - D\vec{\alpha}\|_2^2 \quad s.t. \quad \|\vec{\alpha}\|_0 \leq T, \quad (5.5)$$

where \vec{z} is the vectorized image patch, $\vec{\alpha}$ is the sparse coefficient vector and the constrain on $\|\cdot\|_0$ enforces the sparsity. The output atoms and the corresponding coefficients reflect if the input is blurred and how strong it is.

To be more precise, consider a sharp image patch \vec{x} of size $n \times n$. To represent such a patch using the dictionary D (which contains blur atoms) as accurate as possible and without any restriction on the sparsity, one needs to use a large number of atoms. It can even reach to the patch dimension (n^2). This is due to the fact that, a sharp patch has sharp edges and, in order to decompose it using a dictionary with blurry atoms, a large number of atoms with their corresponding weight coefficients are needed to represent its sharp corners. In contrast, to represent a blurred image patch \vec{y} with the same size, it is possible to get the same accuracy by using fewer atoms in this dictionary. This illustrates that with a fixed sparsity of $T \ll n^2$, we have larger error in representation of a sharp image patch rather than a blurred one. Conclusively, a dictionary containing blur atoms can represent a blur image patch using only few atoms (fixed sparsity) more accurately than representing a sharp image patch with the same number of atoms.

To clarify the concept, a toy example is introduced. Figure 5.4 demonstrates the example. It is shown that having a set of smooth-edged elements in different sizes (a dictionary of blurry atoms, smooth edges), a sharp corner (a sharp image patch, sharp edges) needs more elements from the set to be filled almost completely (to be represented with a low error). However, using the same set (dictionary), a smooth corner (a blurry image patch) can be filled almost completely with less number of elements in the set. This means, if one is allowed to use only i.e. three elements of this set and tries to fill the corners with minimum error, the one with a smooth corner could be filled with higher accuracy rather than the sharp one.

The reason behind using a blur dictionary instead of a sharp one lays on the fact that, in a dictionary with blurry atoms the possibility of having a sharp atom is low. Therefore, decomposition

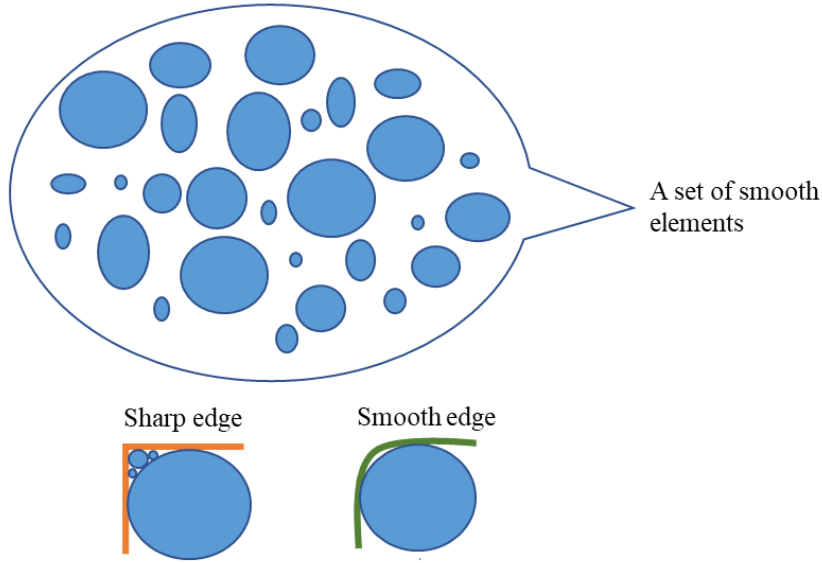


Figure 5.4: A toy example to clarify the logic behind the proposed classifier. Having a set of smooth-edged elements (a dictionary with blur atoms), a sharp corner needs more elements to be filled (coded) with minimum error compared to a smooth corner.

of a sharp image patch with a fixed sparsity leads to a larger error in comparison to the blurry one. In contrast, in a dictionary with sharp atoms, there may exist some flat and smooth regions which are sufficient to represent a blurry image patch with the same sparsity. Thus, a sharp dictionary may have the same ability to represent a sharp and blurry image patch with almost the same error in representation.

Here, the sparsity is an important criterion and the optimal number is defined based on an experimental study. The aim is to have an accurate and sparse coding of signals. It is expected to achieve higher quality (in means of PSNR) using a larger sparsity. However, the point of using sparse representation is to employ a few number of atoms and represent a signal as accurately as possible. This means, the sparsity has to be as small as possible, where using this sparsity one can achieve an acceptable quality (by means of a higher PSNR and lower error in representation). Furthermore, to differentiate between different levels of blurriness using the BM, the smallest sparsity is required. Considering all these criteria, the sparsity 3 has shown to be the best choice.

Thus, the sparsity is set to a fixed integer 3 and the error in representation is calculated. The error in representation reflects if the input is blurry and how strong it is (Figure 5.5, for better visualization of the effect, a bigger patch size is used here.)

$$E = \|\vec{z} - \hat{\vec{z}}\|_2^2 \quad \hat{\vec{z}} = D\vec{\alpha}, \quad \|\vec{\alpha}\|_0 \leq T, \quad T = 3. \quad (5.6)$$

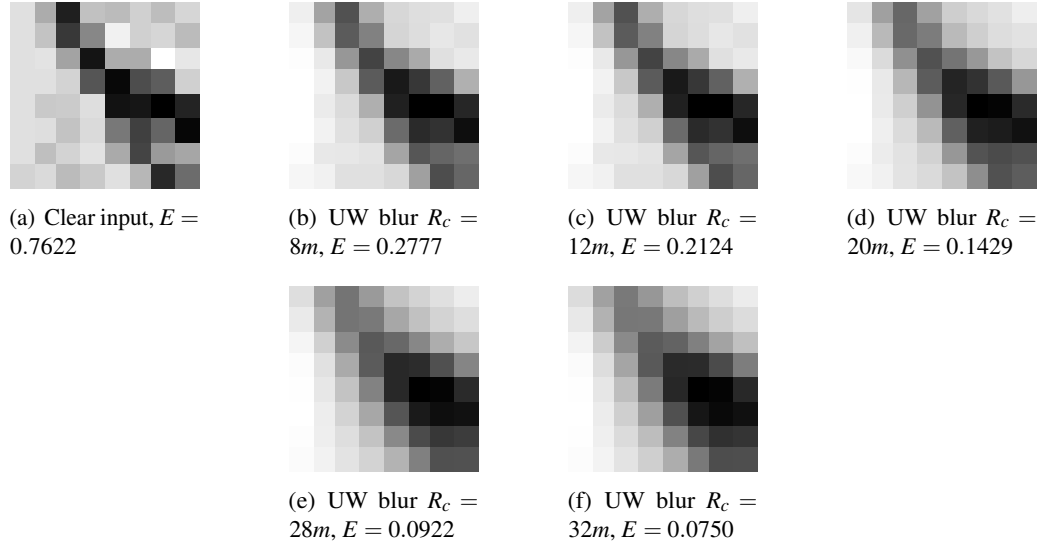


Figure 5.5: Error in representation as a blur measure. As blurriness gets severe (increasing camera-object distance R_c in order of meters), the error in representation using the reference dictionary and forced sparsity 3 decrease correspondingly. Here the water type is considered to be deep ocean with $c = 0.049$.

In summary, a clear image patch with fixed sparsity T , over a dictionary containing blurred atoms, gives a larger error in representation. By increasing the blur level of the image patch, since it matches to the dictionary atoms variation, then the less number of atoms are sufficient to represent it, so the error in representation decreases. The trend can be obtained: larger error in representation, the sharper the image patch is.

5.4 Clustering and Sparse Model Selection with BM

The proposed algorithm is composed of two stages, training and reconstruction. In the first stage, a set of dictionary pairs is prepared and in the second stage the best dictionary pair is selected to sparsely reconstruct HR patches from the corresponding LR counterparts.

5.4.1 Training Phase

The training stage requires a set of *HR* UW images. Then the *LR* version of all those images are constructed by using down-sampling and blurring operators. Bicubic kernel is used to down-sample the images by a factor of 2 and Equation 5.4 is used to simulate UW blur. To reach destination size, *LR* images are interpolated by a scale factor of 2 to the size of *HR* image via Bicubic interpolation. These images are said to be in the middle resolution (MR) level. This scaling serves mainly for rendering the coding part easily.

The main focus in this phase is learning the most appropriate dictionaries to reconstruct edges and texture content of an image accurately. To follow this idea, in order to learn the HR dictionaries from the high components only, the HR images are subtracted from those of the MR ones. Then local patches are extracted and vectorized to form the HR training set X . In order to ensure that the reconstructed HR patch is the best estimation, the calculated coefficients must fit to the most relevant part of its LR signal. Thus the Laplacian and Gradient high-pass filters are employed to remove low frequency content of LR images similar to the approaches in [Yan+08; Yan+10]. This choice is reasonable since human eyes are sensitive to the high frequency component of an image.

Applying high-pass filters on the image patches result in boundary problems due to small patch sizes [IP91]. Therefore, to extract features, the filters are applied on the LR image. As done in [ZEP10], four $1 - D$ filters are applied to extract first and second derivatives:

$$\begin{aligned} f_1 &= [-1, 0, 1] & f_2 &= f_1^T \\ f_3 &= [1, 0, -2, 0, 1] & f_4 &= f_3^T. \end{aligned} \quad (5.7)$$

Then local features corresponding to the gradient maps are extracted and reshaped to the vectors. Features in the same location are concatenated to form a large vector as a feature for the LR training set Y . LR and HR patches corresponding to the same spatial location are handled as pairs.

Once the LR training set, Y , is available, the BM introduced in 5.3 is applied to classify the LR patches. To cluster the patches with rather a similar blurriness level, a set of reference

Algorithm 3 The Proposed Cluster DL Algorithm

Input: HR training image set

Output: A set of HR/LR dictionary pairs, reference patches for each cluster

- 1: Down-sample and blur each HR image to conduct the LR image set.
- 2: Divide each HR and the corresponding LR image into patches.
- 3: Subtract the mean value of each HR patch, reshape them into vectors and concatenate them column-wise to shape an HR training array X .
- 4: Up-sample each LR image to the MR level and apply feature extraction filters on each.
- 5: Divide the extracted feature maps into patches and reshape them into column vectors.
- 6: Concatenate the features column-wise to form the LR training array Y .
- 7: Calculate the BM values of all patches in the LR image (extracted in step 2).
- 8: Cluster BM values using reference patches.
- 9: Add the corresponding MR patch $y \in Y$ into the LR training set of that cluster Y_m .
- 10: Add the pair HR patch $\vec{x} \in X$ to the HR training set of the same cluster X_m .

for each $Y_m \quad \forall m \in \text{number of clusters}$ **do**

Learn corresponding LR dictionary $\{D_l\}_m$ via Equation 5.8.

Calculate the pair HR dictionary for that cluster $\{D_h\}_m$ via Equation 5.9.

end for

patches are employed. Each reference patch corresponds to one level of blurriness (from sharp to a high level of blurriness). They are conducted by applying the blur model 5.4 on sharp and high-quality available patches extracted from UW images. Every LR patch and its corresponding HR counterpart are then clustered accordingly. At the end, for every cluster m , the features from the LR training sets Y_m are given to K-SVD algorithm and the corresponding LR dictionary is learned $\{D_l\}_m \in R^{n \times M}$:

$$\underset{\{D_l\}_m, \{\alpha_k\}_m}{\operatorname{argmin}} \sum_k \|\{y_k\}_m - \{D_l\}_m \{\alpha_k\}_m\|^2 \quad s.t. \quad \|\{\alpha_k\}_m\|_0 \leq T \quad \forall k \in Y_m. \quad (5.8)$$

The DL is done iteratively and is terminated when a stopping criterion is met (error or number of iterations). For the first iteration the initial dictionary is conducted by randomly choosing atoms from Y_m .

Once all LR dictionaries corresponding to different clusters are learned, the HR dictionary counterpart for each cluster is calculated as follows:

$$\{D_h\}_m = X_m(A_m)^T (A_m(A_m)^T)^{-1}, \quad (5.9)$$

here A_m denotes the sparse coefficients of cluster m and X_m stands for corresponding HR training cluster. Algorithm 3 outlines the main steps of the training phase.

5.4.2 Reconstruction Phase

Reconstruction phase should conform to the training phase. In that sense, steps reminiscent of the ones in the learning steps is encountered. Given an *LR* image to be reconstructed, it first needs to be rescaled to the same size as the *HR* image. This is done by Bicubic interpolation. Using this MR image, the meaningful features are extracted in exactly the same way that the learning

Algorithm 4 The Proposed Single Image Super-resolution and Deblurring Algorithm

Input: LR input image, cluster dictionary pairs

Output: HR image estimate

- 1: Divide the LR image into overlapping patches.
- 2: Up-sample the LR image to the required resolution level (MR).
- 3: Apply feature extraction filters, divide the extracted features into overlapping patches and reshape them into vectors.

for each patch extracted from the LR image (step 1) **do**

- Calculate the BM values and define the proper cluster number m .
- Sparsely code the corresponding MR patch $y \in Y$ over the LR dictionary of the cluster m , $\{D_l\}_m$.
- Reconstruct the corresponding HR patch using sparse codes of y and corresponding HR cluster dictionary $\{D_h\}_m$ as: $\vec{x} = \{D_h\}_m \vec{\alpha}$.

end for

Merge overlapping estimated HR patches to obtain an HR image.

stage did. The features are divided into patches and are reshaped to form an array Y . To assure local consistency between the reconstructed patches, a certain patch overlap is allowed [Yan+10] (in this dissertation full overlap is used). Then the features are clustered to be recovered using the most suitable dictionary pair.

The BM for each LR patch is obtained and the proper cluster is defined as it was explained in the training phase. The dictionaries belonging to this cluster are used to reconstruct the HR and blur free output image. For this, the patch which belongs to the cluster m is decomposed using the corresponding LR dictionary $\{D_l\}_m$ and sparse coefficients are calculated. Then, by right multiplying the HR cluster dictionary $\{D_h\}_m$ with the sparse coefficients of the MR patch the HR patch is reconstructed. Finally, all the reconstructed HR patches are reshaped into the 2-D form and merged to constitute the HR image estimate. In this merging, each pixel value is obtained from the average of its values in the reconstructed patches that contain it. Algorithm 4 outlines a summary of the proposed reconstruction procedure.

5.5 Experimental Results and Discussion

In this section, the performance of the proposed method is examined and compared to two leading super-resolution algorithms proposed by Zeyde et al. [ZEP10] and Yang et al. [Yan+10]. The algorithms used in the comparison are along the lines of DL, however, they are different in nature. In order to have fair comparisons, care has been taken to ensure that the parameters used in the training and reconstruction stages are, for all algorithms, as close to each other as possible. If a parameter is unique to a specific algorithm, the value suggested by the authors is used. Image super-resolution results for a scale factor of 2 are presented, although, it can be easily modified for other scale factors.

Before evaluating the performance of the proposed method, a few matters concerning the scale-invariance and generality of the BM, and the patch-size and dictionary redundancy are investigated.

5.5.1 Approximate Scale-Invariance and Generality of the Proposed Blur Measure

To verify the generality of the BM which less error in representation with a fixed sparsity corresponds to a stronger blurriness, thousands of example patches (500,000) are tested. The patches are extracted from high-quality UW images and, using UW blur model [TO06], the different blur levels are simulated. The sharp image patches and the blurred versions are then decoded using the learned dictionary in Section BM 5.3. Here, 30 different versions of blur patches are provided, each correspond to a different level of blurriness. Finally, the average error in representation of all patches (for each level of blurriness) is calculated. Figure 5.6 illustrates the error values E and their corresponding parameter R_c (applying different blurriness levels controlled by camera-scene distance R_c). The sharper the patch is (lower blurriness level) the larger the error is obtained. The error values for different blur levels are consistent which manifest the effectiveness of the metric.

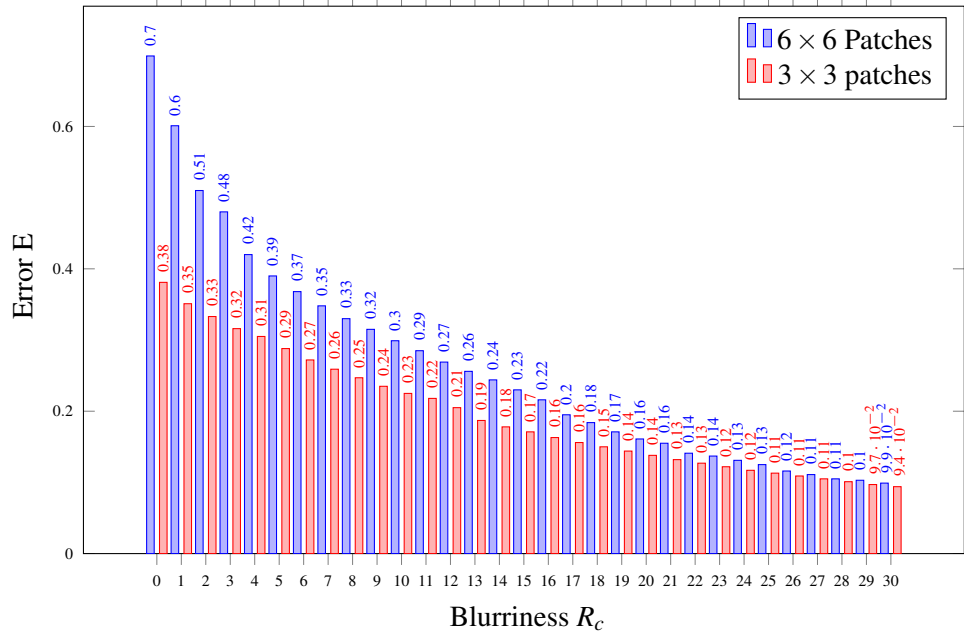


Figure 5.6: Error values vs. blur level. Height of each bar shows the average error in representation of patches with the same blur level using fixed sparsity 3. The chart demonstrates the correlation between the blurriness level and error value for two resolution levels.

To investigate the impact of the scale on the BM, each patch, extracted in the last step, is down-sampled by a scale factor 2 to obtain the LR counterpart. As can be seen in Figure 5.6, the scale does not affect the overall performance of the proposed metric, both HR and LR image patches demonstrate similar distribution. The error in representation reduces as the level of blurriness increases. As a reminder, this is due to coding the patches over a blur dictionary.

Furthermore, a second investigation is carried out by applying the criteria on HR and LR image patches corresponding to some of the test images and the histograms are evaluated. Figure 5.7 shows the corresponding histogram distributions which are approximately similar in both resolution levels (the error values differ which is reasonable due to using two dictionaries with different atom sizes but learned using the same data set). Thus, the statistical distributions of the BM values of the LR and HR image patches suggest that this criterion is approximately scale-invariant.

The above conclusions imply that the BM can be effectively used as a criterion for model selection (dictionary pair) and classification of image patches. In view of Figure 5.6, one can see that the difference between the average MSE value of sharp image patches and blur ones is bigger than the difference between two different blur levels, especially when the patches have higher levels of blurriness. This is logical since there is a higher probability that a blur dictionary corresponds to i.e. level 25, may have a similar ability to represent patches with a higher level of

5.5 Experimental Results and Discussion

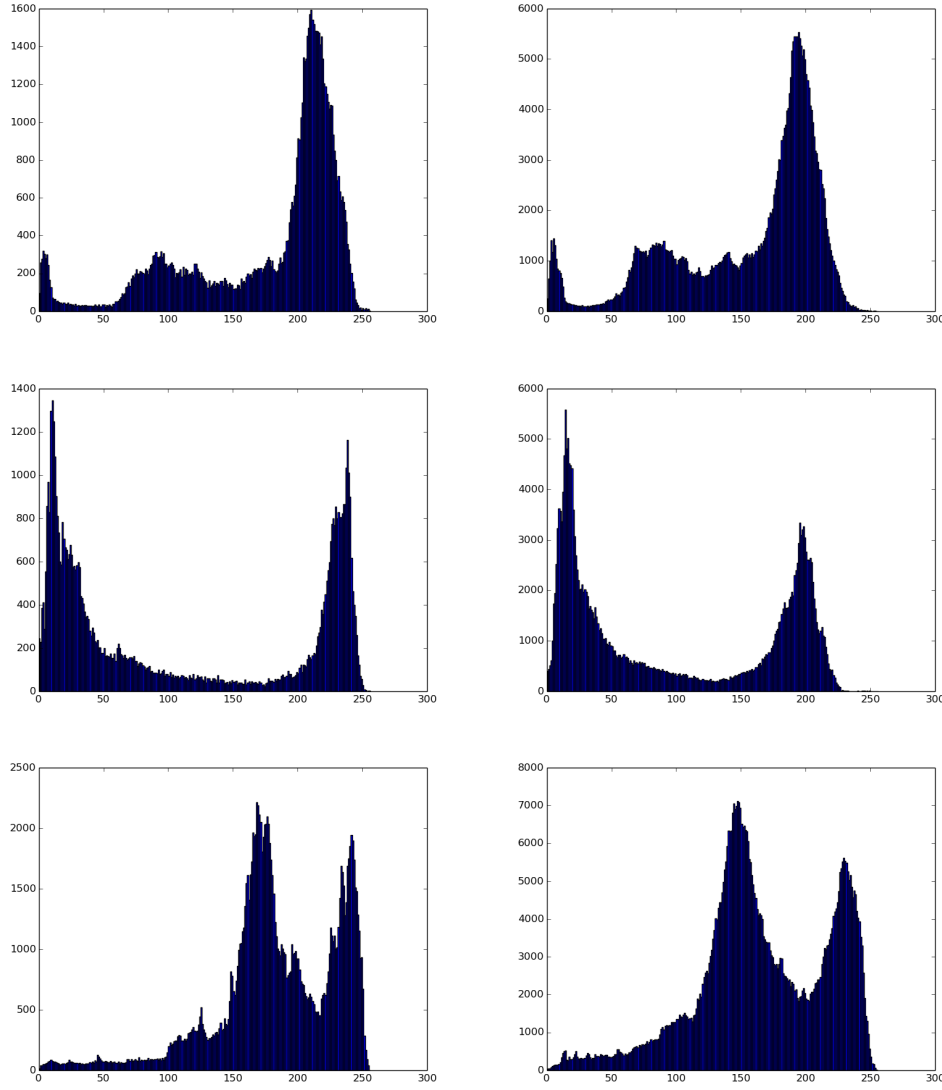


Figure 5.7: Histogram of BM values for HR patches (left), and LR patches (right) for *coral reef 1*, *coral reef 2* and *submarine* images in Figure 5.12, 5.13 and 5.14 respectively.

blurriness (i.e. 26) rather than a dictionary corresponding to i.e. level 5. However, the proposed BM could differentiate between those patches with a very small yet meaningful difference.

The pattern in values of BM eases the clustering of sharp image patches especially. The sharpest patches are those which contain edges, corners and texture, in other words, they are patches of high frequencies and the most difficult image regions to reconstruct. This observation forms the basis for the potential improvement in HR image reconstruction, depending on the availability of cluster dictionary pairs that can effectively represent image patches in their respective clusters.

5.5.2 Effect of the Patch Size and the Dictionary Redundancy on the Representation Quality

According to [BDE09] dictionary redundancy has an important role in the sparse representation problem. This problem has a patch-based approach, therefore, the redundancy of the dictionary is defined by the ratio of the number of dictionary atoms to the patch size. Thus, it is essential to identify the best choice for the patch size and number of dictionary atoms to achieve a high-quality in the restored image.

It is expected that a larger patch size is more effective in representing the image structures, such as edges. At the same time, this leads to larger dictionary atoms which means a significantly large data set is needed. Furthermore, the complexity is another concern which will be considerably high in case of large dictionary atoms. To this end, a study which shows the ability of designed dictionaries using different settings for the patch size and the dictionary redundancy is carried out. Then, using the optimal patch size and dictionary redundancy, where both a high-quality and a low complexity are fulfilled, the evaluation of the proposed method is performed.

The steps are as follows: first, several dictionaries having different parameters, patch size and dictionary redundancy, are learned. The dictionaries are learned using KSVD algorithm [AEB06] with sparsity $T = 3$ and 25 iterations. Next, a test image data set is sparsely coded over each of the learned dictionaries. Finally, the average PSNR of each category, corresponding to different dictionaries with different settings, is calculated.

Figure 5.8 illustrates the result of this study. The horizontal axis represents the redundancy as the ratio of number of dictionary atoms (columns) to the patch size (the number of dictionary's rows), and vertical axis stands for PSNR performance of the coded data set. The study has been derived for four HR patch sizes 4×4 , 5×5 , 6×6 and 10×10 which correspond to the 2×2 , 3×3 , 4×4 and 5×5 LR patch sizes respectively.

The training set (images are used for learning dictionaries) are not used in the coding step, meaning the testing and the training data sets are independent.

In view of figure 5.8, one can see that increasing the dictionary redundancy improves the results for all patch sizes. This is logical since the more atoms in a dictionary means the more options to find the best ones to represent different patches, especially when the representation is sparse, meaning, only a few atoms are allowed to be used for the best possible representation. Thus, using an over-complete dictionary has a great impact on the final results rather than a complete-dictionary.

The main goal is choosing the best dictionary redundancy which improves the result while keeping the computational complexity reasonably low. The study shows that except HR patch size 4×4 which does not provide good results, the rest of the patch sizes have a rather similar ability to represent the data set (the difference between PSNR values is rather small). As can be observed, the patch size 6×6 is the best choice with the highest PSNR values. This is in

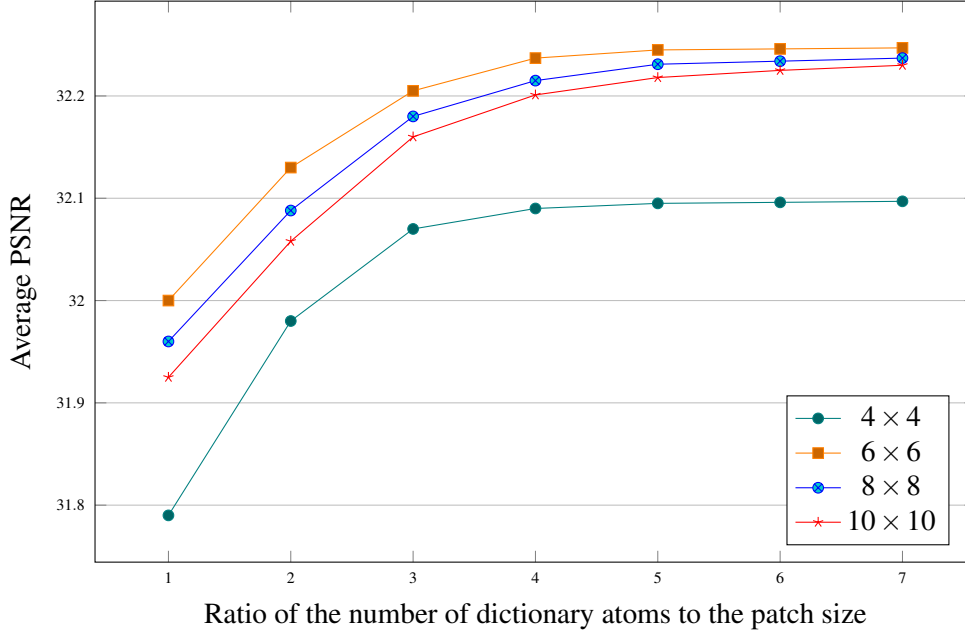


Figure 5.8: Average PSNR vs. dictionary redundancy for four different HR patch size 4×4 , 6×6 , 8×8 and 10×10 correspond to LR patch size 2×2 , 3×3 , 4×4 and 5×5 respectively.

contrast to the expectation that larger patch sizes are able to represent the image structures better. Generally, this is a true statement, however, when the representation is done over a dictionary with a limited number of atoms and a fixed sparsity then the scenario differs. According to the Figure 5.8, using a patch size large enough to represent the simple image structures (i.e. horizontal, vertical and diagonal edges), in this study 3×3 , one can achieve a good PSNR value. Although, when the patch size is larger, 4×4 and 5×5 , the more complicated image structures are extracted (i.e. curves, mixture of simple edges). This means, one needs to have a very rich dictionary (many atoms covering all combinations of simple structures) and a higher sparsity to find the best combination of atoms to represent them as good as a simple image structure. This theory is proved in this study, as it can be seen, when the dictionary redundancy increases the PSNR values of bigger patch sizes (4×4 and 5×5) are improved. At the higher redundancies, there is almost no difference between the PSNR value of these image patch sizes and the 3×3 image patch size.

The last observation is that, using image patch size 3×3 , increasing the dictionary redundancy more than 4 does not improve the PSNR values much. Therefore, the LR patch size 3×3 and a dictionary redundancy of 4 seems to be a good compromise in terms of performance and computational complexity.



Figure 5.9: Examples of the training images.

5.5.3 Implementation Details and Evaluations

According to the study above, a LR patch size of 3×3 is used for both learning and reconstruction phases. The dictionaries are learned with redundancy 4 which leads to the dictionaries of size 36×144 . For training, one pixel overlap is used where a full overlap (two pixels) is used in the reconstruction phase. The reason is that in the training step there is no need to extract all the possible patches since there will be many similar patches, which is not helpful. However, for the testing stage it is beneficial to extract all the possible patches since the overlapping parts can correct any probable mismatch.

Dictionaries of the proposed method are trained in a coupled manner as outlined in the training phase (Section 5.4.1). DL is done with over 500,000 pairs of LR and HR patches randomly selected from the training set. Some of the images in the training set are shown in Fig. 5.9. The images are chosen based on their rich high frequency components.

The LR test images are obtained by down-sampling the HR input images by 2 in both dimensions and then the forward scattering blur filter explained in 5.4 is applied on this LR counterpart. For the evaluation of the proposed method, a set of LR natural UW images are employed. These images suffer from blur degradation due to forward scatter component and have a low resolution (mostly 338×234). A LR grayscale image is the input to the proposed algorithm to reconstruct its HR counterpart. Thus, color images are first transformed to the luminance and chrominance

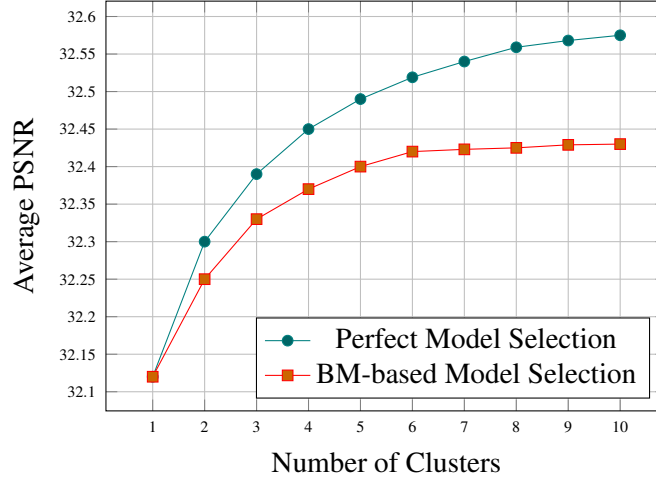


Figure 5.10: Performance of the proposed algorithm (the average PSNR versus number of clusters) using the perfect model selection and the BM as a model selection criterion.

color space and only the luminance component is the input to the algorithm to reconstruct the luminance component of the corresponding HR image. The two other color channels are then resized to the destination size by applying the Bicubic interpolation. Finally, all the channels are merged and the image is converted back to RGB color format.

The other two algorithms, [ZEP10; Yan+10], are set to have the parameters which are suggested by the authors, a LR patch size of 3×3 with 2 pixels overlap and dictionary of size 36×1000 for [ZEP10] and a LR patch size of 5×5 with 4 pixels overlap and dictionary size 100×1025 for [Yan+10].

An important issue in the proposed algorithm is the number of clusters to employ. There is no specific rule of thumb to decide on the optimal number of clusters. Thus, the impact of number of clusters on the performance is empirically studied. Using BM and LR patches, the training data is classified in 2, 3, 4, 5, 6, 7, 8, 9, 10 clusters. For each case, the dictionary pairs are learned. In addition, a single cluster is also taken into account which learns a single coupled dictionary pair. Using the BM criterion, all the test images are clustered, coded and the average PSNR is calculated. Furthermore, the same experiment is run using perfect model selection, meaning, all the patches are coded using every cluster and the one which minimizes the MSE value is chosen as the perfect dictionary pair. The plots are illustrated in Figure 5.10.

Several observations can be derived from the results in this Figure:

- It indicates that using a set of cluster dictionary pairs is better than employing only a single dictionary pair. The reason is clear since the forward scatter blur is spatial variant thus, one dictionary with a fixed level of blurriness can not properly restore UW images which are degraded by this kind of blur. This is proved by the perfect model selection plot which indicates the more clusters lead to a better performance.

Table 5.1: MSE performance of the dictionary pairs at restoring HR image patches in different clusters (the blurriness increases by the index of clusters, lowest blurriness: C_1 and highest blurriness: C_6). A single dictionary pair (C_{all} , learned over all training patches as one cluster) is also evaluated.

Data in / Dictionaries in	C_1	C_2	C_3	C_4	C_5	C_6	C_{all}
C_1	301.194	353.356	388.760	395.091	409.190	414.104	391.001
C_2	143.091	127.452	151.001	167.921	169.240	173.010	158.065
C_3	79.034	72.743	63.089	69.284	73.090	75.419	69040
C_4	44.606	41.044	35.129	32.892	39.010	36.024	36.561
C_5	21.302	18.984	16.090	15.452	13.079	14.001	16.201
C_6	3.509	3.890	3.350	2.209	2.006	1.832	2.879

- It shows that using the BM for classifying patches and the model selection, improves the performance of the algorithm by increasing the number of clusters.
- The average PSNR shows a significant improvement until the 6th cluster and afterward does change slightly. This shows a trade-off between the accuracy of the BM as a model selection and the number of clusters.

As a conclusion, the BM as a model selection criterion is able to discriminate between patches with different blurriness levels and therefore, represent them better. Yet, using more clusters is not very helpful since it does not improve the PSNR values much but increases the complexity. It can be explained, since the BM selects the best dictionary pair based on the ability of a blur dictionary at representing image patches, so, the higher the patch blurriness is, the smaller the difference between the error in representations (Figure 5.6). Thus, more clusters means that these small differences are critical which may end up to some misclassifications. In addition, it is possible that each dictionary contains some atoms with a higher blurriness (especially when the difference between levels of blurriness is not very high), meaning they may be able to represent some patches with higher level of blurriness as good.

Conclusively, the result of this study suggests that 6 number of clusters is a good bargain between accuracy of the model selection, PSNR performance and complexity.

Once the number of clusters is set, the ability of the BM at the dictionary selection step is investigated. For this, the training set of each cluster is used. The LR image patches in each cluster are restored using all the dictionary pairs and the MSE between the reconstructed version and the original HR patches are calculated. The approximated image patches corresponding to each cluster (C_1, \dots, C_6) should give the minimum average MSE. This is due to the fact that the dictionary pair in each cluster should have the highest ability to restore the LR training patches of the same cluster.

Table 5.1 depicts the outcomes. It clearly indicates that the results fulfill the expectation. In addition, it can be observed that almost every dictionary pair can restore the blur patches

successfully (especially C_6 and C_5 with highest blurriness). As the level of blurriness decreases (especially the C_1 with lowest blurriness), the image patches cannot be represented by other dictionary pairs satisfactorily. Besides, one can see that the image patches with lowest blurriness have the highest MSE which indicates the challenge of reconstructing the sharp image features in general.

Figure 5.11 demonstrates the learned HR dictionaries for all 6 clusters where the atoms of each dictionary inherit the blurriness nature of their cluster. The blur level of the atoms increases from D_1 to D_6 .

Using multiple dictionaries, it is expected that the computational complexity increases. It is well-known that the most computationally complex stage is the sparse coding step. Since the cluster dictionary pairs of the proposed method are learned off-line, thus, the most computation-

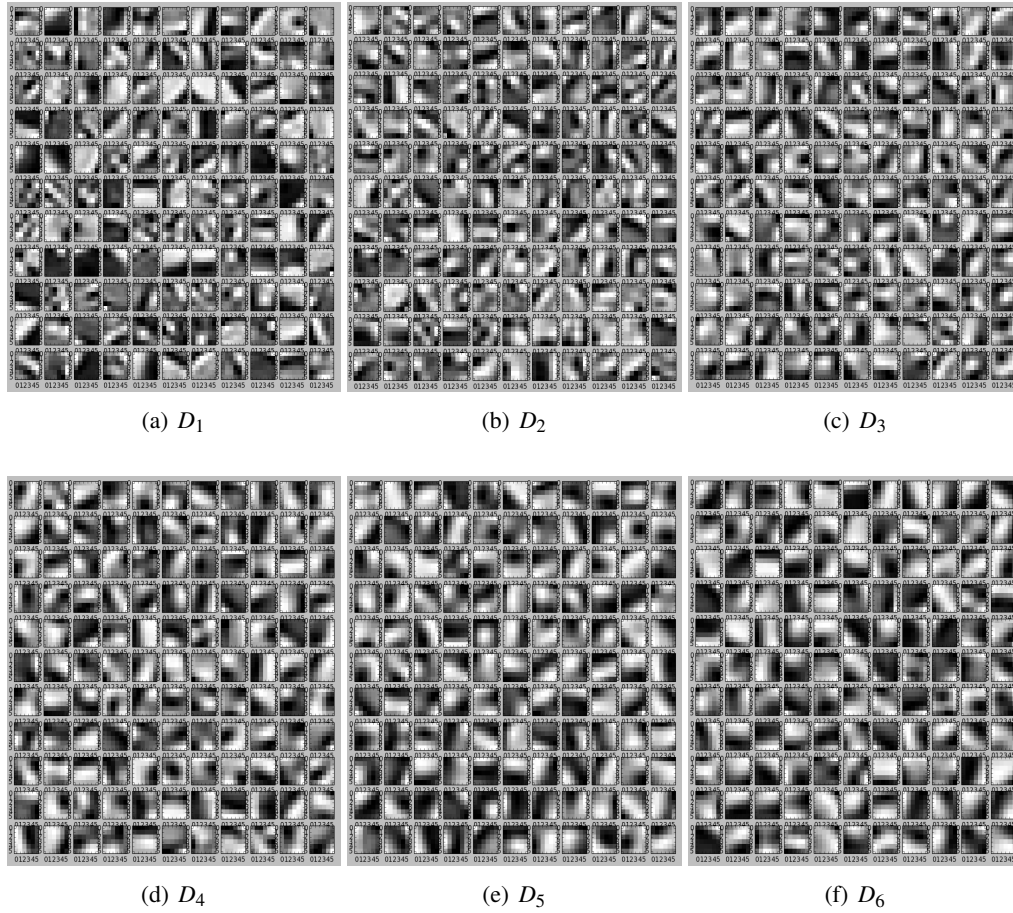


Figure 5.11: The learned HR dictionaries corresponding to 6 different clusters. The blur level of the atoms of the dictionaries increases from D_1 to D_6 .

ally expensive stage in this frame-work is the sparse coding in the reconstruction phase which relies on vector selection [ROF92]. However, only one pair is selected for each patch where the dictionary pair has a low redundancy. Therefore, computational complexity for each cluster is much lower than that in a single high redundant dictionary with 1000 atoms in [ZEP10]. As a result, using six compact dictionary pairs of size 144 even reduces the overall SR computational complexity.

In order to quantitatively evaluate the performance of an algorithm, it is common practice to use PSNR, MSE and the structural similarity index measure (SSIM). However, all of these metrics are reference-based and it is necessary to access the ground truth image. For UW images, usually, there is no such possibility. Thereby, non-reference quality metrics are used to evaluate the quality of the results.

UW color image quality evaluation (UCIQE) [YS15] and UW image quality measure (UIQM) [PGA16a] are two non-reference quality metrics which are proposed specifically for UW images. According to the degradations present in UW images, color cast, low contrast and blur are considered. Thus, three values measuring the UW image colorfulness (UICM), sharpness (UISM) and contrast (UIConM) are provided. Here, since the proposed algorithm deals with low resolution and blur rather than color cast, the UISM seems to be a reasonable choice for the performance evaluation. In addition, in situations where contrast is degraded and therefore, edges are not easily detected (i.e. images in Figure 5.14 and 5.15), the proposed algorithm enhances the local contrast. Thus, two measures of this metric, UISM and UIConM, are considered as evaluation metrics in this chapter.

The UISM deals with the sharpness of the image edges, thus, the edge map of each color channel is obtained and the enhancement measure estimation (EME) [PSA14] is applied on each edge map as follows:

$$EME = \frac{2}{k_1 k_2} \sum_{l=1}^{k_1} \sum_{k=1}^{k_2} \log\left(\frac{I_{max,k,l}}{I_{min,k,l}}\right), \quad (5.10)$$

where the image is divided into $k_1 k_2$ blocks and $(I_{max,k,l}/I_{min,k,l})$ indicates the relative contrast ratio within each block. Once the EME is calculated for each color channel, the UISM is computed via a linear combination of them,

$$UISM = \sum_{\lambda=1}^3 \gamma_{\lambda} EME(\text{grayscale edge}_{\lambda}), \quad (5.11)$$

here $\lambda \in RGB$ and $\gamma_R = 0.299$, $\gamma_G = 0.587$ and $\gamma_B = 0.114$. To obtain the grayscale edge map, first, the Canny edge detector is applied on each color channel and the result is multiplied with the original image. A higher value of UISM indicates a better deblurring result.

On the other hand, UIConM [PGA16a] is computed via logarithmic average Michelson contrast measure by entropy (logAMEE) [Pan+11] in local image regions as follows:

$$UIConM = \logAMEE(\text{intensity}) \quad (5.12)$$

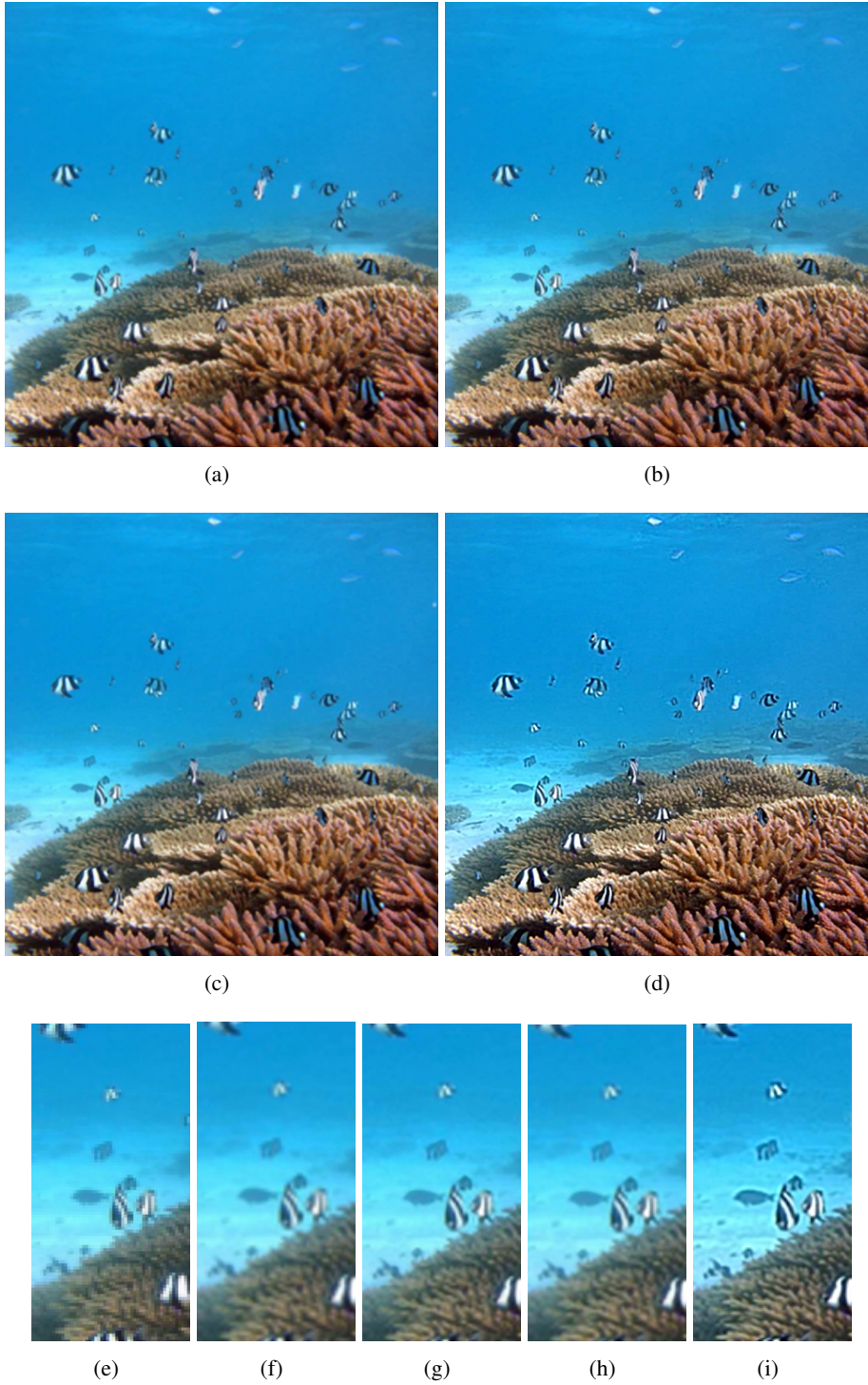


Figure 5.12: Visual comparison of the *coral reef 1* image: reconstructions using (a) Bicubic interpolation, (b) Yang et al. [Yan+10], (c) Zeyde et al. [ZEP10], and (d) the proposed algorithm. The last row, (f-i) show zoom-in presentations of each algorithm respectively which are compared to the LR input image (e).

where

$$\log AMEE = \frac{1}{k_1 k_2} \otimes \sum_{l=1}^{k_1} \sum_{k=1}^{k_2} \frac{I_{max,k,l} \ominus I_{min,k,l}}{I_{max,k,l} \oplus I_{min,k,l}} \times \log \left(\frac{I_{max,k,l} \ominus I_{min,k,l}}{I_{max,k,l} \oplus I_{min,k,l}} \right) \quad (5.13)$$

where the intensity image is divided into $k_1 k_2$ blocks, and \ominus , \oplus and \otimes are PLIP operations [Pan+11]. PLIP operations provide nonlinear representations that are consistent with human visual perceptions (for more on this subject, the reader is referred to [Pan+11]). A larger value stands for a better-contrasted image.

Following other researchers, the entropy of the output image is also provided. Entropy is a statistical measure and provides a value indicating the amount of information used to characterize the texture of an image. It is computed as:

$$\text{Entropy} = - \sum_{k=0}^{M-1} p_k \log_2(p_k) \quad (5.14)$$

here M is the number of gray levels and p_k is the probability associated with the gray level k . The higher entropy means more information is provided by the result.

The result of the proposed algorithm is compared to two leading super-resolution algorithms, [ZEP10; Yan+10], plus a Bicubic interpolation to illustrate the difference.

Figures 5.12, 5.13, 5.14 and 5.15 give a qualitative comparison between the proposed algorithm and three aforementioned approaches. The images in Figures 5.12, 5.13, 5.14 and 5.15 are named *submarine*, *coral reef 1*, *coral reef 2* and *statue* respectively. The selected images are the ones containing objects in different camera-scene distances to examine the capability of the proposed algorithm for handling spatial variant blur.

Figures conduct a comparison between the Bicubic interpolation (a), the Zeyde et al. algorithm [ZEP10] (b), the Yang et al. approach [Yan+10] (c) and finally the proposed super-resolution and deblurring algorithm (d). (f-i) provide a zoom-in presentation of each algorithm respectively which are compared to the up-sampled LR input image (e).

In view of Figures 5.12, one can observe that simply zooming the LR image leads to blockish artifact (e), where a simple interpolation approach such as Bicubic (f) can overcome this problem. However, it causes blurriness. A super-resolution algorithm can avoid the blockish effect and restore the image structures while decreasing the blurriness introduced by interpolation algorithms. It can be seen that, compared to Bicubic interpolation, the approaches in [ZEP10] and [Yan+10] increased the image resolution when the edges are preserved and are sharper. The result of the proposed method suggests a superior improvement over the other three algorithms. It did not only increase the resolution of the image without bringing blur but also sharpened the image structures and details significantly. All the objects in the scene, despite their distance to the camera, are enhanced. The details on the coral reefs at the far distance are only visible in the result of the proposed method. This is where there are no over/under-sharpening of the edges which could result in halo artifacts or the loss of image details.

The same comparison is applied on the *coral reef 2* image (Figure 5.13). The same as the previous image, the details and image structures in the output of the proposed approach are

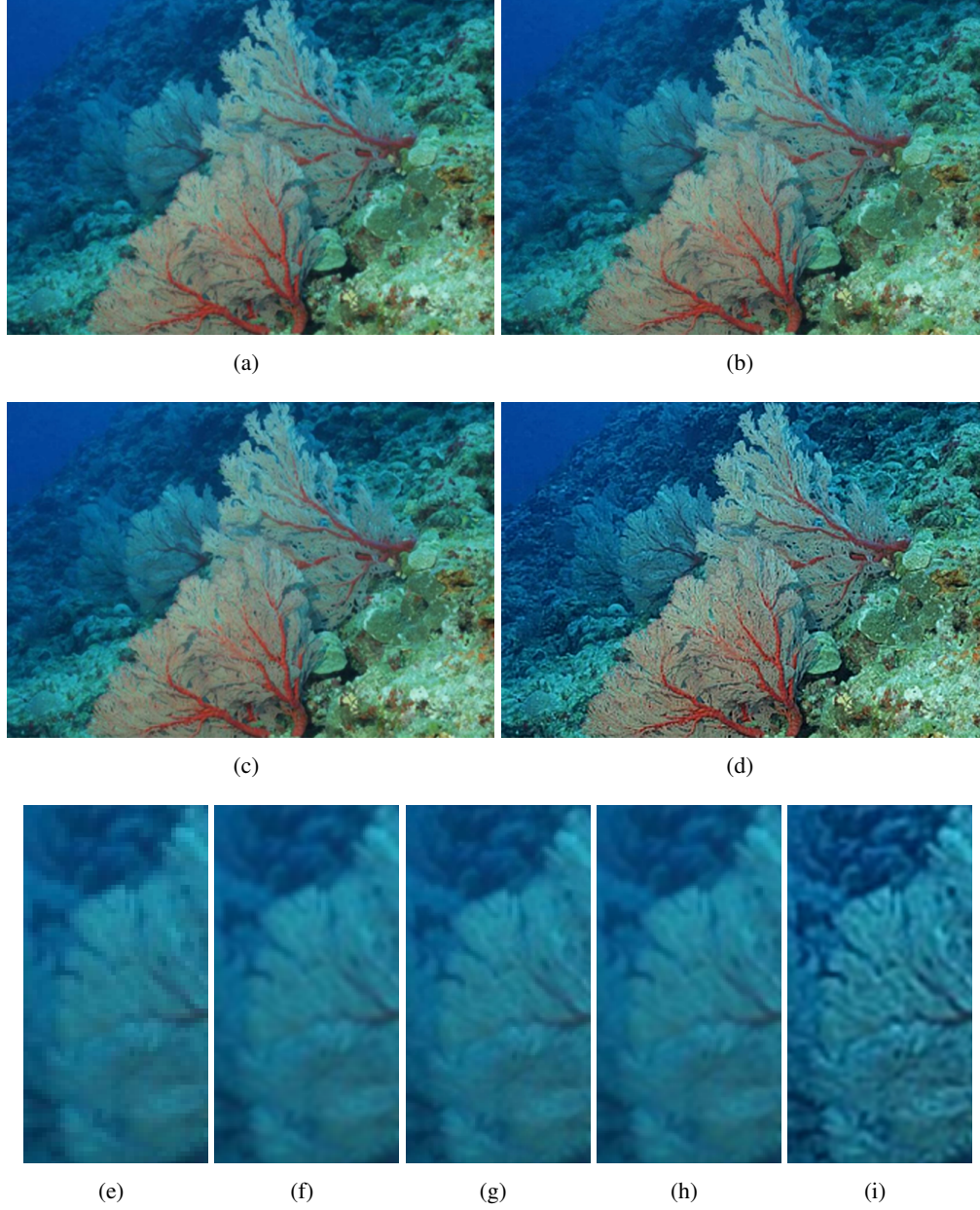


Figure 5.13: Visual comparison of the *coral reef 2* image: reconstructions using (a) Bicubic interpolation, (b) Yang et al. [Yan+10], (c) Zeyde et al. [ZEP10], and (d) the proposed algorithm. The last row, (f-i) show zoom-in presentations of each algorithm respectively which are compared to the LR input image (e).

enhanced clearly. The edge between the structure and water (left-up) is very well defined. The same effect can be seen on the small details of coral reefs. The output image obviously provides a high resolution and much more clarity of image structures compared to state-of-the-art approaches.

The proposed algorithm is evaluated using *submarine* and *statue* images (Figure 5.14 and 5.15 respectively) which are not only blur but also low-contrasted. In these cases, small edges and structures are degraded further due to slight haze which adds an extra challenge for the proposed method. Despite the extra difficulty, as can be seen in Figures 5.14 and 5.15, the proposed algorithm could restore those small details very well resulting in a better local contrast besides higher resolution and deblurring. In both Figures, the proposed algorithm improved image resolution and sharpened image structures in different camera-scene distances where no artifacts are introduced. The zoom-in presentations clearly illustrate the improvement obtained over the leading algorithms.

The qualitative evaluation of the proposed algorithm is in line with the quantitative result presented in Table 5.2. This table demonstrates the quantitative presentation of images in Figures 5.12, 5.13, 5.14 and 5.15 compared to Bicubic interpolation, [ZEP10], and [Yan+10]. The evaluation is done based on image sharpness (UISM), contrast (UIConM) and entropy.

In view of this table, one can see that the proposed algorithm has achieved the highest image sharpness compared to the other algorithms. The Bicubic interpolation caused image blurriness, therefore, its UISM value decreased compared to the original image. The purpose of the super-resolution algorithm is to increase the image resolution when the image features are restored as accurately as possible. It can be seen in results of the other two algorithms which super-resolved the LR input image and decreased the blurriness compared to the Bicubic interpolation (UISM values are almost competitive to the original image, especially in case of Yang et al. algorithm [Yan+10]). The proposed algorithm achieved the highest UISM values suggesting super-resolution without introducing blur, but also sharpening image details (enhance the original blurry input image).

The values of entropy support this conclusion where the highest values belong to the proposed method. The entropy of an image is an indicator of the amount of information it offers. In all four images, as can be visually observed, the proposed algorithm enhances the image details and structures which leads to more available information.

The contrast values show the improvement, especially for the *submarine* and *statue* images. This is due to the presence of haze in these images which causes lower contrast on the image edges. The proposed algorithm improves the contrast in the local regions by enhancing the image sharpness resulting in higher UIConM value.

As a by-product, the proposed BM can be used to define an approximate distance map. Since the BM is an indicator for image patch blurriness and also considering the fact that UW blur varies according to the camera-object distance, thus, it can be used to define a distance map of the objects in the scene. Figure 5.16 illustrates the extracted map corresponding to the *coral reef 1*, *coral reef 2*, *submarine* and *statue* images. It can be observed that the distance maps have a patch-wise accuracy, this is due to the fact that the BM defines a value for each local

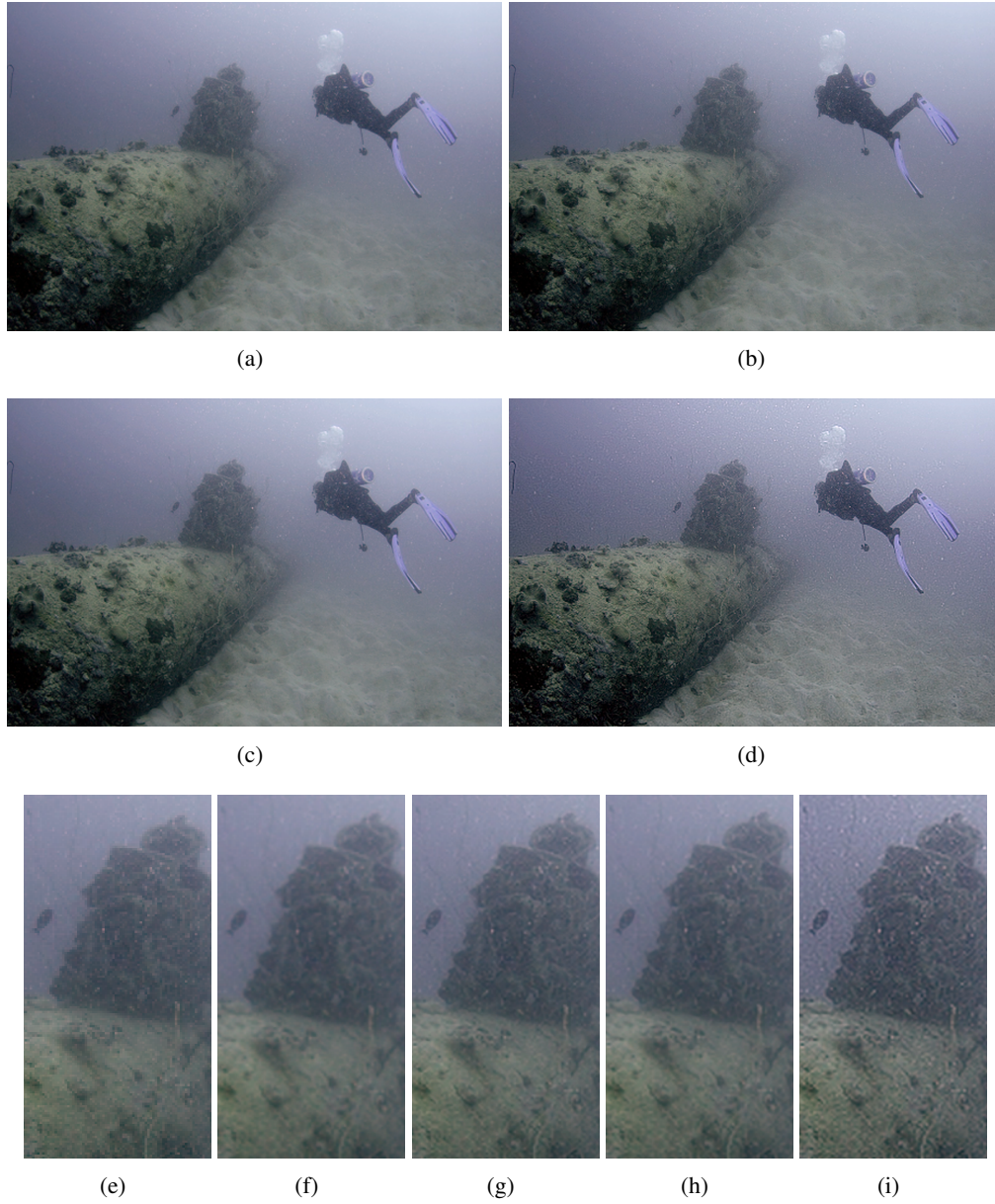


Figure 5.14: Visual comparison of the *submarine* image: reconstructions using (a) Bicubic interpolation, (b) Yang et al. [Yan+10], (c) Zeyde et al. [ZEP10], and (d) the proposed algorithm. The last row, (f-i) show zoom-in presentations of each algorithm respectively which are compared to the LR input image (e).

patch. However, one can refine the map and get a pixel-wise accuracy by applying algorithms such as image matting [LLW08]. The distance map can be used in several scenarios such as 3D reconstruction of the scene and image segmentation.

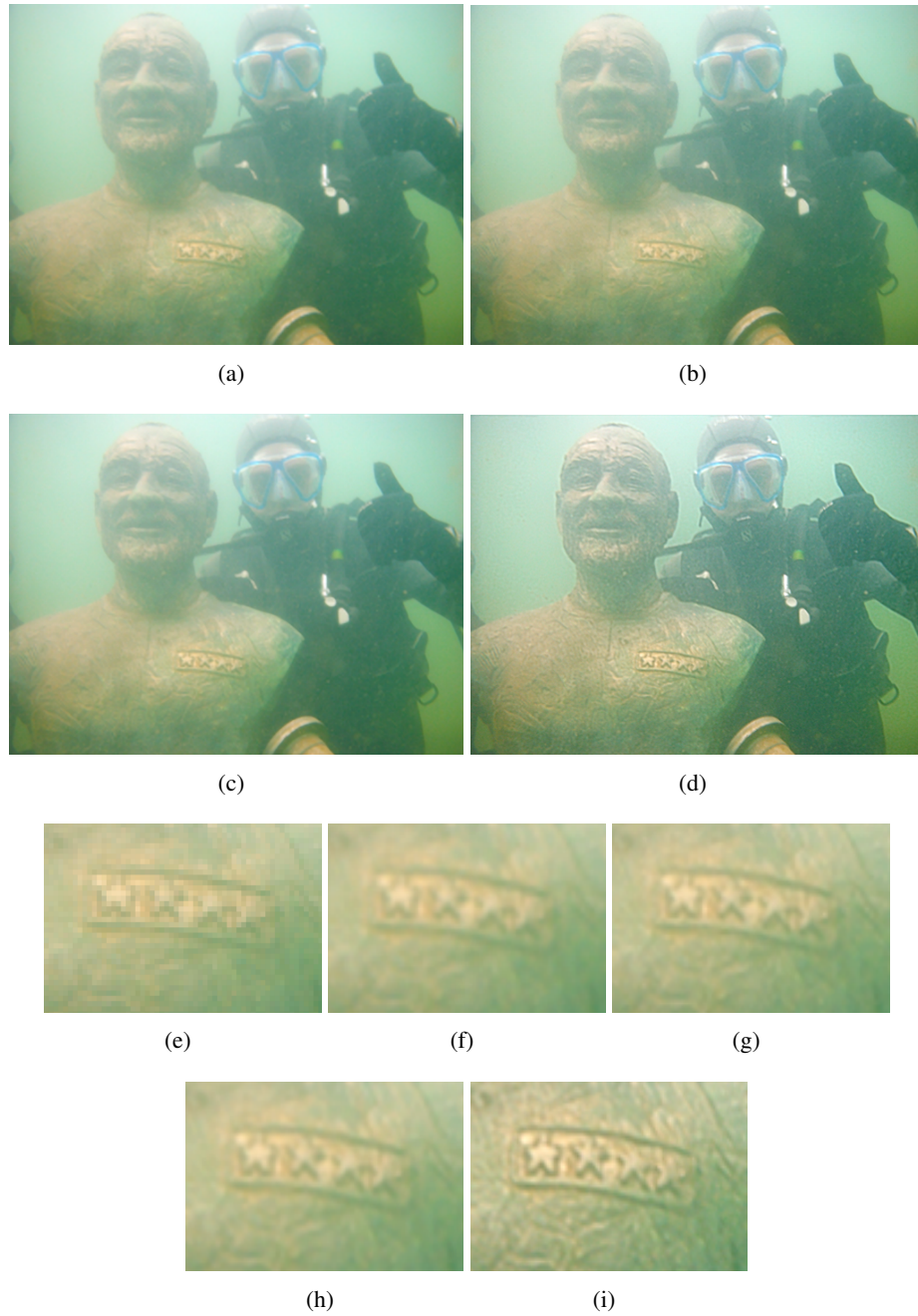


Figure 5.15: Visual comparison of the *statue* image: reconstructions using (a) Bicubic interpolation, (b) Yang et al. [Yan+10], (c) Zeyde et al. [ZEP10], and (d) the proposed algorithm. The last row, (f-i) show zoom-in presentations of each algorithm respectively which are compared to the LR input image (e).

Table 5.2: Quantitative comparisons (UISM, UIConM and entropy) of the reconstruction using Bicubic interpolation, Zeyde et al. [ZEP10] algorithm, Yang et al. [Yan+10] algorithm and the proposed approach.

		Original	Bicubic	[ZEP10]	[Yan+10]	Proposed Method
UISM	submarine	2.2054	1.4700	1.7075	1.8510	2.8387
	coral reef 1	3.6394	3.2136	3.5297	3.6760	5.4359
	coral reef 2	3.4160	2.9647	3.2433	3.3151	5.8217
	statue	0.8070	0.6746	0.7065	0.7487	1.2080
UIConM	submarine	0.1124	0.0959	0.1008	0.1051	0.1268
	coral reef 1	0.0848	0.0857	0.0838	0.0856	0.0856
	coral reef 2	0.1395	0.1382	0.1389	0.1393	0.1395
	statue	0.1012	0.0906	0.0925	0.0969	0.1207
Entropy	submarine	7.0226	7.0306	7.0338	7.0368	7.0844
	coral reef 1	6.2166	6.1796	6.2071	6.2199	6.4200
	coral reef 2	6.8110	6.7931	6.8073	6.8091	6.9915
	statue	7.0170	7.0089	7.0116	7.0130	7.0498

Yet, it should be considered that the extracted distance map is an estimation and the accuracy of it may decrease if the image is degraded with strong haze or color cast where the image structures are faded.

5.6 Conclusion

This chapter sought to improve UW image resolution over a set of compact class dictionaries. For this, a classifier, which could perform clustering of training and testing signals, was needed. such a classifier is used to describe a certain signal class, meaning, the signals belonging to a certain class share a common property. Since the super-resolution is concerned, another requirement imposed upon the classifier is the scale-invariance, meaning the decision made about clustering of a patch into a certain cluster should be as independent of scale as possible.

To this end, a blur measure (BM) based on sparse representation and dictionary learning was developed. The measure is based on the capability of a blur dictionary for coding sharp and blur patches. The studies showed that the proposed BM is approximately scale-invariant and holds for different levels of blurriness.

In addition, an image enhancement algorithm was integrated into a super-resolution approach which provides sharper image structures (deblurring). Before applying the proposed algorithm, the effect of different parameters, such as patch size, dictionary redundancy and number of clusters were investigated and the optimal values considering the quality performance and complexity were defined.

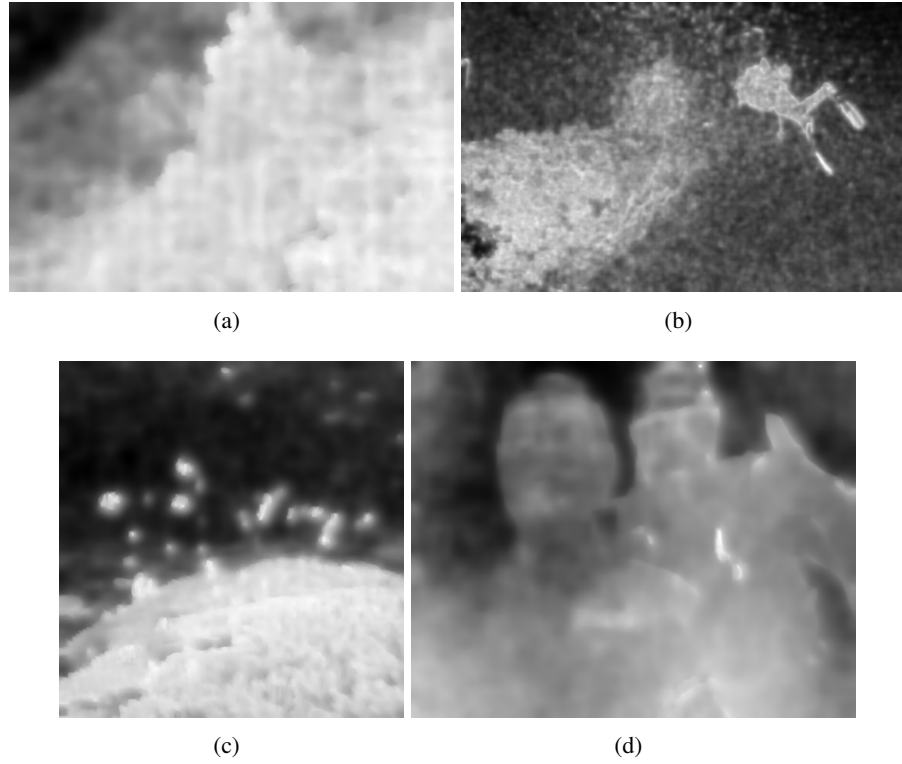


Figure 5.16: Examples of estimated distance map using BM. (a-d) belong to the *coral reef 1*, *submarine*, *coral reef 2*, and *statue* images.

When applied to the UW images, the proposed sparse coding paradigm was shown to produce the state-of-the-art performance. The quantitative and qualitative evaluations suggested that the proposed algorithm not only super-resolves the LR images but also sharpens image structures. This is especially advantageous for UW images which suffer from spatial variant blur.

However, like many other algorithms, the proposed algorithm has shortcomings too. The BM clusters image patches by measuring the blurriness of the structures such as edges, yet, when the image is highly corrupted by haze, it is possible that edges and details are faded, therefore, not readable for the algorithm. This can interrupt the performance of the algorithm and lead to inaccurate model selection.

In conclusion, the following points were made through the work conducted in this chapter.

- The proposed BM is used to separate image patches based on their spatial intensity variations. It is used for clustering and model selection for the purpose of super-resolution.
- Super-resolution and deblurring via sparse representation over multiple clusters where a pair of dictionaries is trained for each cluster, is shown to produce results that are superior to the results of the-state-of-the-art super-resolution algorithms ([ZEP10; Yan+10]).

- The proposed algorithm uses several dictionary pairs, however, each dictionary is compact. Thus, its computational complexity is comparable to the standard sparse coding.

The results of the super-resolution and deblurring algorithm conducted very promising results where even very small image details are taken care of. As an example, in the *submarine* image (Figure 5.14) there are many small particles which are not clearly visible in the original image. However, after restoration, one can observe them easily. These floating particles, containing dead material and organic matters, are very common in UW images. Their existence, however, reduces the scene visibility. Therefore, they are mostly considered as an unwanted signal in an image. In addition, image processing and computer vision tasks (i.e. segmentation and tracking) are affected by the presence of these particles.

As a result, it is advantageous to detect and eliminate these particles from UW images as a pre-processing step for advanced image processing and computer vision tasks. To this end, the elimination of floating particles, so-called *marine snow*, is considered. In the next chapter, two approaches dealing with marine snow elimination in single images and video frames are addressed.

6 Denoising: Marine Snow Detection and Removal

6.1 Introduction

It is a common problem that images captured UW are corrupted by noise. The major barrier is that light, unlike sound, is poorly propagated in water. This is explained by the propagation properties of light in water ([McG80; Wel69]). Light is exponentially attenuated while traveling in water, which is due to the light absorption and scattering by the marine environment. Therefore, visibility in this medium is limited up to few meters or even less, depending on water type (i.e. deep ocean, coastal and bay).

Despite blur, haze, low contrast, non-uniform lightening and color cast, which are occasionally termed noise, additive noises, such as sensor noise, are the main objective for denoising algorithms. However, visibility of UW scenes is distorted by another source termed "marine snow". This signal not only decreases scene visibility by its presence but also disturbs the performance of advanced image processing algorithms such as segmentation, classification or detection.

Despite the fact that, in orders of magnitude, marine snow together with backscatter, has the greatest degradation factor [AMK05b], it is not well-researched and mostly neglected from image processing algorithms. Minimal amount of literature have addressed the elimination of such a phenomenon; in most cases, marine snow is clustered in the same group as additive noises with a general approach being proposed by the authors. Thus, UW images are considered to be polluted by basic noises such as salt & pepper, Gaussian or speckle noise. Examples of such algorithms are filtering [AMK05a; Liu+15b], wavelet decomposition and high-pass filtering [FXG11; PK10], a combination of curvelet and filtering [SSS13].

However, the main assumptions of these algorithms (additive and single pixel) do not match characteristics of marine snow. Marine snow is part of a scene which consists of particles with sizes larger than a single pixel. Thus, one cannot treat it the same way as additive noises; it needs to be addressed directly.

In this chapter, a new perspective towards UW image denoising is adopted. Marine snow is considered as the main objective and appropriate approaches specifically addressing elimination of these particles are proposed. To detect any distinct object, one needs to study its properties; hereby, the characteristics of marine snow in images/frames are investigated. Next, the proper actions are taken to detect and eliminate it. Two different approaches are suggested, one addresses marine snow removal in single images where an estimation of the scene behind marine snow is

obtained using neighborhoods. For this, a novel approach, termed Median² filtering, is proposed. It detects the particles and filters them in a new fashion which can successfully:

- filter the particles with a structure of several pixels at once,
- distinguish between the object edges and the particles,
- reduce the blurriness effect introduced by filtering.

The second approach deals with video frames where the camera is static. For this scenario, image restoration is applied where successive frames are used to learn image priors and the scene behind marine snow is recovered employing inpainting. For this, the frames of a video are divided into two groups of training and testing. The training frames are employed to learn a background model of the scene without marine snow. This image providing information about the scene covered by marine snow is then used to learn the statistics of the whole scene. In the meantime, for each testing frame a mask defining the locations of particles is obtained. The proposed Median² approach is used to extract such a mask without filtering the original image. Once the mask is available, the scene behind particles is restored using the priors without using the neighborhood pixels directly. Thus, it

- achieves higher accuracy in terms of image statistics,
- further reduces the blurriness effect,
- takes care of the temporal coherence of video since the background model is used as the training data for learning priors.

Furthermore, the approach is adapted to the situations where the camera has small movements such as minor pan, tilt or zoom. This is especially beneficial when a robot is used to acquire the data since the slow movement of the robot can be considered as one of these movements. Thus, the proposed approach can also be used for such scenarios.

The rest of this chapter is organized as follows: Section 6.2 gives a short introduction about marine snow and its definition. Section 6.3 reviews the related work. Single image marine snow detection and removal is then proposed in Section 6.4 where the elimination of marine snow is applied via image filtering. In Section 6.5 a second approach suited for videos with a static camera, where small movements are allowed, has been proposed. Each algorithm is evaluated qualitatively and quantitatively and lastly, a short conclusion will conclude the chapter.

6.2 Marine Snow

Much of the suspended matter in the ocean exists as aggregates of organic detritus, microorganisms, and clay minerals. These particles range from a few microns to many centimeters in size and are abundant and every-where throughout the world's oceans. Marine snow is herein defined as those aggregated particles in the ocean larger than 500 μm in length. The name "marine snow"

comes from an early *in situ* observation of particles undertaken by Japanese researchers in the 1950s [SK53].

These particles sink through a primary mechanism by which surface derived material reaches deep water and the ocean floor. Through traveling deeper and deeper, the particles can grow in size as a smaller flakes clump. The larger sizes fall more quickly through the water column but, even so, the journey to the bottom can take several weeks. Not all particles get that far since they are mostly a source of food and are often consumed by fish or marine mammals during their slow descent. Once the journey is complete, this decomposing mess is a reliable food source on the sparse darkness of the sea floor.

The structure of aggregates varies along a continuum from fragile, porous, loose associations of smaller particles and organisms to highly cohesive, robust, gelatinous structures produced by zooplankton. This is due to its highly diverse origins, morphologies and characteristics. Shapes range from compact spheres to comets, strands and plates.

Despite their highly variable physical characteristics, appearance, and composition, marine snow is formed from only two general pathways [AS88]:

- Aggregates which are produced by mucus-producing marine organisms, especially by particular zooplankton and phytoplankton;
- Aggregates which result from the biologically-enhanced physical aggregation of smaller component particles.

For more on this topic the reader is referred to [AS88].

6.2.1 Marine Snow as a Source of Noise and its Characteristics

Despite the importance of marine snow for marine researchers and their significant role as a source of food, these particles can be considered as a source of noise for image processing and computer vision tasks.

In an image, marine snow appears as white bright spots, which can degrade the scene visibility where the objects of interest are something other than the particles. Figure 6.1 illustrates some example UW images where the particles degrade the scene visibility and are in some cases, misleading.

Furthermore, the performance of image processing algorithms is affected by the presence of these particles. Such a scenario can be seen in segmentation, detection or tracking algorithms where marine snow leads to misclassification or false detection. Since marine snow is a true object in the scene, applying image processing algorithms such as enhancement, super-resolution or color correction also enhances the marine snow. This could be considered as enhancing noise which is, however, undesirable. Figure 6.2 shows three example images before and after applying deblurring, color correction and segmentation algorithms.

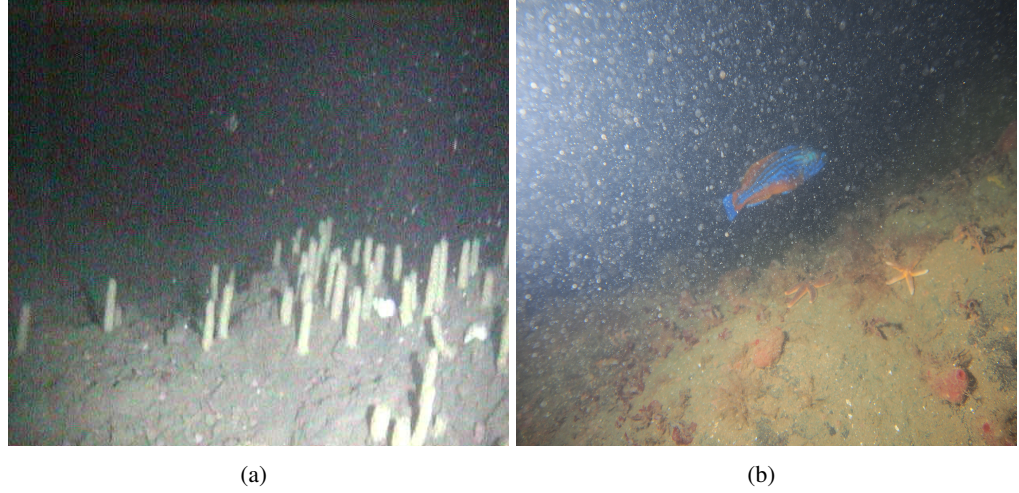


Figure 6.1: Examples of images distorted by marine snow: (a) marine snow causes misleading where this frame, together with many others, is used to restore a 3D model of the scene, (b) image visibility is limited due to a heavy marine snow.

Generally speaking, image noise is an unwanted signal which usually appears as random speckles (variation of brightness or color information) on an otherwise smooth surface and can significantly degrade image quality. Considering the difficulties caused by the presence of marine snow and the fact that it is an unwanted signal for many image processing and computer vision tasks, here it is termed noise, thus, one needs to eliminate it before applying any advance computer vision algorithm. To this end, first, characteristics of marine snow in an image are investigated and second, two approaches are proposed accordingly. The main goal is to remove marine snow when even small edges are preserved.

Characteristics of Marine Snow

Since marine snow is not a typical noise but instead a true object in the scene, the existing denoising approaches do not succeed in removing this phenomenon. In order to propose an approach which detects and removes marine snow, one needs to know what characteristics describe it and how it appears in an image.

In order to be able to detect the particles, the physical characteristics of marine snow in many UW images and video frames are investigated. The outcome is listed as follows:

1. Size: it appears in different sizes depending on the image resolution and the camera scene distance. In the images provided in this work, where the resolution is mostly 1920×1080 , it is between 3 by 3 pixels to 25 by 25 pixels (Figure 6.3(a)).
2. Geometry: it can be roughly estimated having Gaussian distribution in all directions, but not necessarily all Gaussians are the same size. (Figure 6.3(b)).

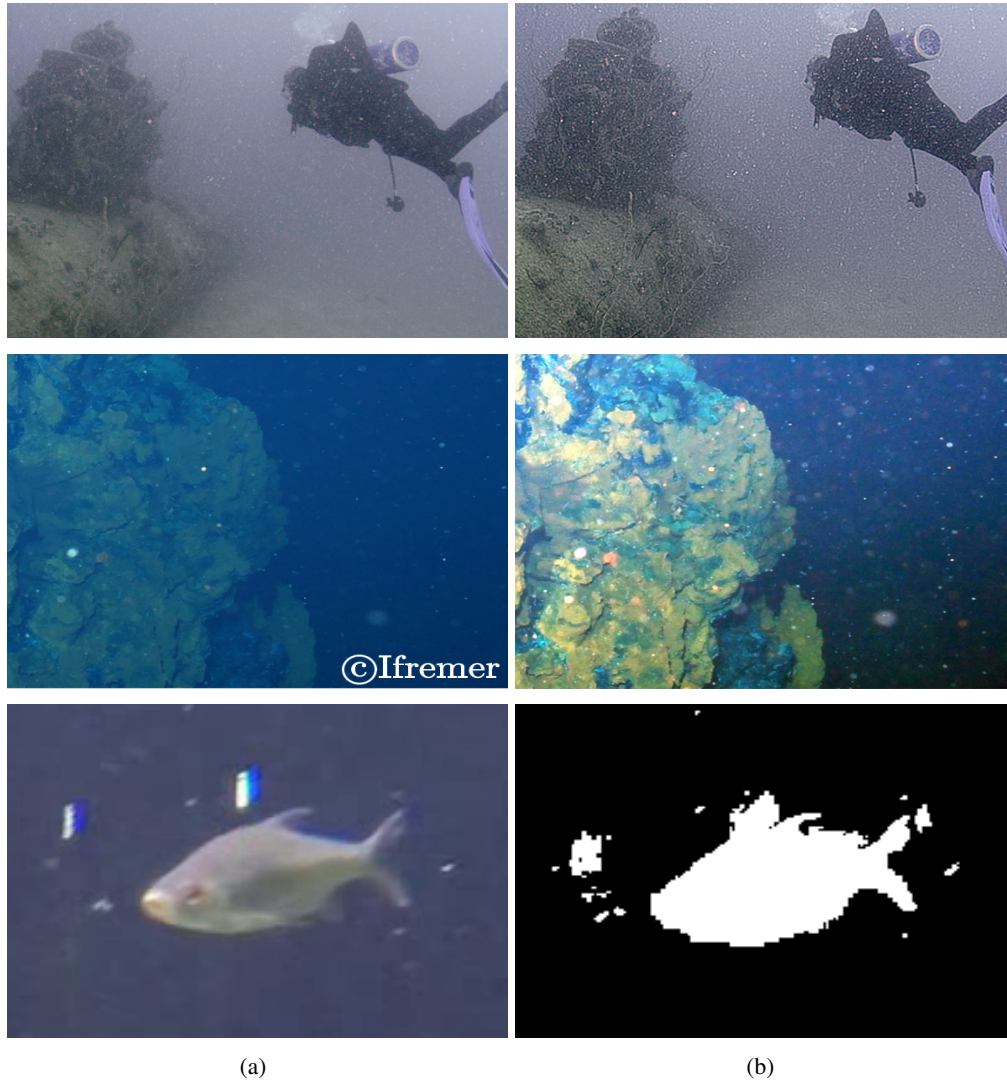


Figure 6.2: Examples of images distorted by marine snow (a), and the results after applying two different image processing algorithms and a computer vision approach (b). First row: deblurring algorithm, second row: color correction approach and the third row: segmentation. In image processing scenarios (the first two rows) marine snow is enhanced the same as other objects in the image, where the segmentation (last row) is affected by the presence of marine particles. Image in the second row is courtesy of Boffety and Galland [BG12].

3. Intensity: it consists of a high peak in the middle and the intensity in the surrounding area decreases proportional to the distance to the peak location. In most cases, a dark area around the marine snow highlights its intensity (Figure 6.3(b)).

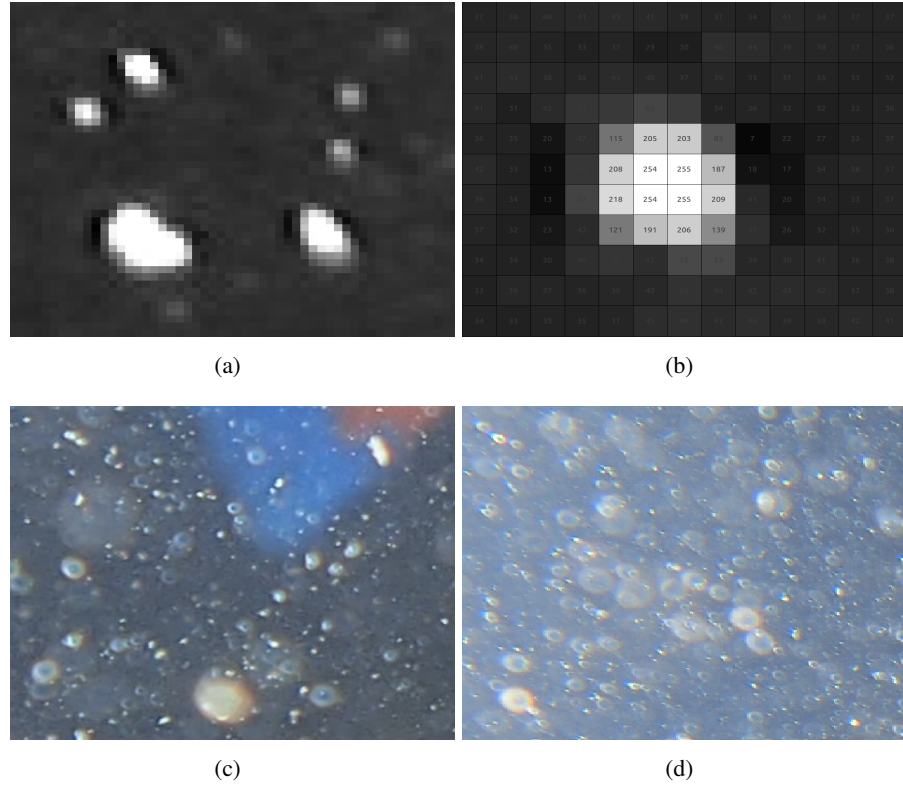


Figure 6.3: Illustration of marine snow characteristics. (a) different sizes and geometries, (b) the intensity distribution of an example particle, (c) overlapping particles, and (d) the reflection effect of particles due to strong backscattering in the presence of an artificial light.

4. **Overlap:** consider a scene divided into many layers (each layer consists of the objects in a specific camera-object distance). Thus, in contrast to an additive noise, marine snow can be present in all layers of a scene. This means that, in an image, the particles can have a highly overlapped distribution (Figure 6.3(c)).
5. **Reflection:** the most challenging fact about this phenomenon is that, especially, in the case of an artificial light, it scatters the light to the camera and appears as circle shaped reflections. This leads to further distortion since the reflections superimpose themselves in front of the scene (Figure 6.3(d)).

Figure 6.4 shows two images with marine snow where an artificial light is used. It can be observed that the upper-left side of both images contain a strong light reflection (circle-shape reflections of marine snow). This is due to the positioning of the artificial light. Comparing the left and right side of the images, one can inspect the effect of different light conditions in the presence of marine snow.

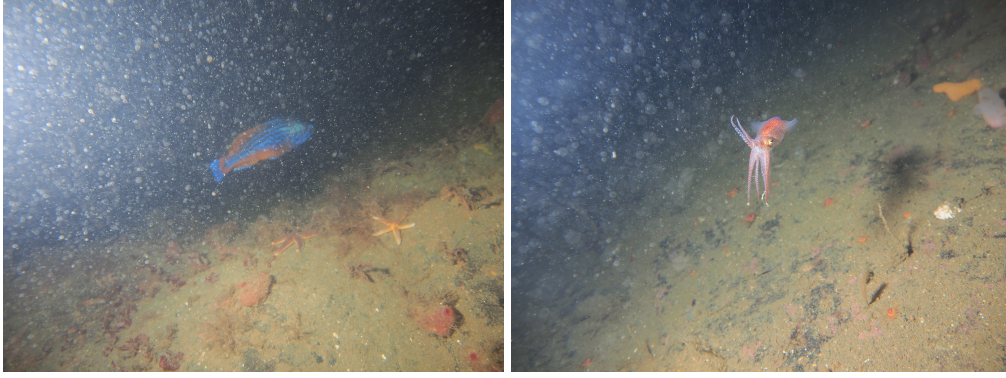


Figure 6.4: Example of images with marine snow where an artificial light is used. The comparison of the left and right sides of both images illustrates the effect of backscattering on marine snow in presence of artificial light.

6.3 Related Works

With knowing the characteristics of marine snow, the next step is to detect it in an image and then to remove it (restore the scene behind the particles). Once the available data is single images, the best approximation of the scene behind marine snow can be obtained using neighborhood pixels.

A simple and low complex alternative is Median filtering. Examples of successful approaches employing this filter can be seen in dust and scratch removal [Ber+08] and varieties of Median filtering approaches tackling impulse noise [WZ99; HH95; CW01; CHN04]. The main objective of this filter is to clean the image from the occurrence of unwanted signals where the edges are preserved as much as possible. The fundamental difference between these approaches is the strategy of detecting the noise. Once the noise or defect is detected, they employ the basic Median filtering to eliminate it. The following provides a short summary of some of the successful approaches using Median filtering.

The basic Median filtering is a technique which a square window of size $2k + 1$, where k goes from 1 to N , is used to filter the center pixel by replacing it with a median value.

There are more advanced techniques based on Median filtering such as the progressive Median filter [WZ99], and decision based Median filter [SE07]. These approaches are designed in two steps to detect and remove impulse noise. Wang and Zhang [WZ99] defined the corrupted pixels in a specified window/patch Ω as follows:

$$|p^{(n-1)} - \text{med}[\Omega^{(n-1)}]| > T_D \quad p \in \Omega, \quad (6.1)$$

where, $p^{(n-1)}$ is the image pixel value in Ω and iteration $n - 1$, and T is a predefined threshold. Using this, the value of each pixel is then modified as (to be used for next iteration):

$$p^{(n)} = \begin{cases} \text{median}[\Omega^{(n-1)}], & \text{if } f^{(n)} \neq f^{(n-1)} \\ p^{(n-1)}, & \text{if } f^{(n)} = f^{(n-1)} \end{cases}, \quad (6.2)$$

here $f^{(n)}$ and $f^{(n-1)}$ are the flags of the current pixel at iteration n and $(n-1)$ respectively. The flag defines if a pixel is corrupted with value 1 or 0 if it is noise free. Once the algorithm detects corrupted pixels, then in the second step, it filters the noisy pixels in each window by replacing them with the median of uncorrupted pixel values in that window. This step is repeated until all pixels are flagged as noise free.

[ZLL07; SE07] follow almost the same logic with different selection criteria. At the first stage, the algorithm checks if the median value of the current window is an impulse or not. For this they use a logic as follows:

$$\min(\Omega) < \text{med}(\Omega) < \max(\Omega), \quad (6.3)$$

where $\min(\Omega)$, $\text{med}(\Omega)$ and $\max(\Omega)$ are the minimum, median and maximum value in a specified window Ω respectively. The window size is increased until either the condition is fulfilled or the maximum window size is reached. Once the median value is a non-impulse value, the algorithm follows the second level. There the current pixel value is verified for being noisy. If it is corrupted then the value of this pixel is replaced by the median value or else it will remain unchanged.

Lastly, Banerjee et al. [Ban+14] proposed a probabilistic approach using Median filtering to eliminate marine snow from single images. This approach is determined as the only available literature which directly addresses marine snow removal. It checks the probability of the existence of marine snow in each patch. This is done by looking for high-luminance pixels within a patch using a predefined threshold and calculating its probability as follows:

$$P(MS) = 1 - \frac{N_{HL}}{N}, \quad (6.4)$$

where N_{HL} and N stand for the number of high luminance pixels and a total number of pixels in the current patch respectively. They consider a cross-checking step to avoid misclassification of the true objects as marine snow. To this end, keeping the same center pixel, they increased the patch size by 2 (in both directions) and calculated the probability one more time. If the probability of having marine snow in the resized patch is still high (low number of high luminance pixels) then the center pixel is replaced by the median value of the local patch. The logic behind this is that they assumed marine snow to have a structure of two or three pixels; therefore, if the probability of high luminance pixels increased it means that it is a larger object which is then assumed not to be marine snow. This algorithm could eliminate small particles, however, it suffers from several shortcomings as follows:

- They assumed that marine snow is structured as two or three pixels. However, this assumption does not always hold true, since usually marine snow, depending on the image resolution, has larger structures (in our case it reaches 25×25 pixels). Thus, considering it to have sizes larger than three pixels, this criterion cannot differentiate between marine snow and other objects in the image.
- Simply increasing the patch size to take into account larger sizes of marine snow may lead to a significantly blurred image.

- It does not use all the information provided in an image since it only considered grayscale images which could therefore result in false detection of similar structures with different colors.

To overcome the aforementioned drawbacks, when the simplicity and efficiency of Median filtering is also taken into account, a single image marine snow detection and removal is proposed. The proposed algorithm follows the same basic idea with aforesaid approaches, where Median filtering is used to estimate the noise free scene but differs at detection of corrupted pixels and the filtering scheme. The following section provides a detailed explanation of the proposed algorithm.

6.4 Single Image Marine Snow Removal via Median² Filtering

To eliminate marine snow from single images, the first step is to detect the pixels corrupted by the particles. This requires an approach which not only finds the location of particles, but also detects the whole structure. According to the properties of marine snow, each particle has a structure of several pixels, between 3×3 to 25×25 , and the intensity changes through the structure. If the whole structure of the particles is undetected, it may affect the estimation of pixel values behind marine snow. Therefore, precise detection is required. The second step deals with filtering the corrupted pixels. Since the structures of the particles are larger than a single pixel, simple Median filtering will result in blurring the image structures. Thus, a voting algorithm is considered to assist the filtering step which reduces the blurring effect considerably.

The procedure for the detection and removal of marine snow is as follows: first, the algorithm looks for the potential corrupted pixels with marine particles. The candidate pixels are the ones which are dissimilar from their neighbors (by means of intensity) and, therefore, are suspected to be marine snow. This is due to the appearance of marine snow, a sudden high intensity occurrence in a local patch. In this stage, the set of detected pixels will be large and it may happen to contain many false detections of the edges of the objects. This is done to make sure that as many pixels as possible that are corrupted with marine snow are detected. However, it is necessary to limit the set by differentiating between the edges of the objects and marine snow.

To do so, the candidate pixels are further limited based on their saturation, and density. This leads to discarding the candidate pixels with vivid colors and the edges of the large structures (the details are provided in the next two subsections). The pixels in the set are then filtered by replacing their value with the corresponding median value of the current patch. The median value of each patch is calculated from the remaining pixel values.

Before the final result is obtained, cross-checking is applied. In this stage, the set of the corrupted pixels is decreased once more by discarding the edges of the small objects. The proposed algorithm extracts the patches on a highly overlapped basis, meaning each pixel is represented in n^2 patches except for those pixels on the border of the image (where $n \times n$ is the patch size). Thus, each suspected pixel is explored in all possible patches. In this stage, a voting algorithm makes the final decision. If, in the majority of the candidate patches, the suspected

pixel is defined as marine snow, then the value of this pixel is replaced by Median² value. This value is calculated from the median of median values of the corresponding patches. Otherwise, it indicates that the pixel belongs to an object edge and therefore, retains its original value. The motivation for cross-checking is to prevent the algorithm from smoothing the image structures due to filtering plus differentiating between the small edges of the objects and marine snow.

To be more precise, consider the corrupted image Y of size $N \times M$ where $Y_\lambda(i, j)$ denotes the luminance value at pixel location (i, j) for each channel $\lambda \in RGB$. Let Ω be the extracted patch of size $n \times n$ centered at $Y_\lambda(i, j)$. The details of the proposed algorithm are explained in the following two subsections.

6.4.1 Coarse Detection

Here, a rough detection of corrupted pixels, that is pixels containing marine snow, within the patches of an image, is obtained. This is done based on the characteristics of marine snow, as listed in Section 6.2.1.

The first observation, that one can derive from the appearance of marine snow in an image, is its intensity. Generally, in an image marine snow is represented as bright spots, thereby, a sudden high-intensity occurrence in a local patch is marked as potential marine snow. A candidate pixel p has to satisfy the following inequality:

$$\|p - \mu(\Omega)\|_2^2 > C \cdot \sigma(\Omega) \quad p \in \Omega, \quad (6.5)$$

here C is a constant controlling the effect of image statistics, $\sigma(\Omega)$ is the standard deviation and $\mu(\Omega)$ represents the mean value of the local patch Ω .

The next step is to identify the general outliers within the candidate pixels of the last step. This is done to distinguish between a high-intensity outlier and a high-intensity object edge. For this, the idea of Gutzeit et al. [Gut+10] is employed. The RGB color space is considered as Euclidean space. Pixels within the current patch are then represented in this space and the density surrounding each candidate pixel is then calculated. Figure 6.5 demonstrates this process. A sphere covering an area surrounding each high-intensity pixel, defined in the last step, is explored. The number of pixels within this sphere

$$\#\{v \in \Omega \mid \exists p \in \Omega : \|p - v\|_2^2 < C \cdot \sigma(\Omega)\}, \quad (6.6)$$

together with the volume of the sphere and the overall number of pixels in the patch provides the density. To make the approach adaptive, the radius of the sphere is calculated dynamically. It is based on the weighted standard deviation of local patch Ω .

A further observation concerning marine snow can be obtained according to its colors. It mostly appears as having high-intensity and low saturation. Thus, by applying the following inequality the pixels with high saturation are discarded:

$$|p_c - p_l| < T \quad \forall c, l \in \{R, G, B\} \wedge c \neq l. \quad (6.7)$$

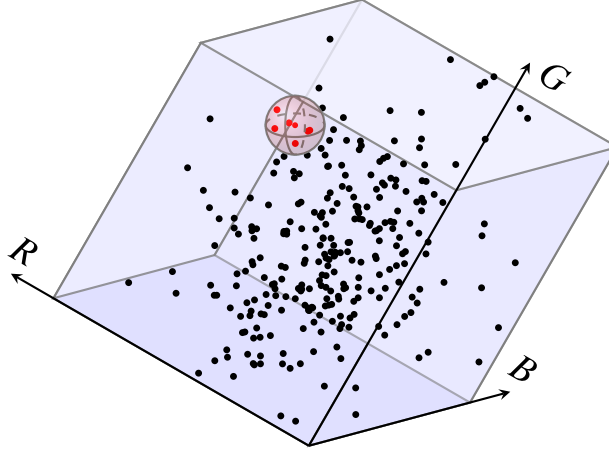


Figure 6.5: The pixels of current patch are visualized as points in RGB color space. The red sphere demonstrates the search environment for the density calculation.

Thereby, the candidate pixels are limited to have colors close to white using a predefined threshold T (based on experimental analysis, $T = 0.1p$ is chosen where p is the intensity value of the current pixel).

All the pixel values who satisfy the aforementioned conditions are then discarded and the median value of the remaining pixel values within local patch Ω is calculated. For now, all the eliminated values in this patch are replaced by this median value:

$$\Omega_f(i, j) = \begin{cases} \text{med}(\text{RPV}), & \text{if } \Omega(i, j) \in \text{DPV} \\ \Omega(i, j), & \text{otherwise} \end{cases} \quad (6.8)$$

where Ω_f indicates the filtered counterpart of patch Ω , RPV stands for *remaining pixel values* and DPV denotes *discarded pixel values*. This procedure is repeated for the whole image.

Before filtering, the pixels surrounding DPV are examined for a sudden low intensity; these pixel values are also discarded from the median calculation. The reason for such a selection is to discard the halo-like shadows around the high intensity of marine snow (see property 3 of Section 6.2.1). Therefore, not only a very high bright spot but also a very dark area surrounding it, defines this phenomenon and should be filtered out.

6.4.2 Fine Detection and Filtering

To do further double checking, to avoid removing the objectedges instead of marine snow, a voting algorithm has been developed. It is responsible for cross-checking all the selected pixel values in a patch to ensure against false detection.

To this end, the patches are extracted on a highly overlapped basis; this means each pixel can be in n^2 possible patches except for the pixels located on the border of the image ($n \times n$ is

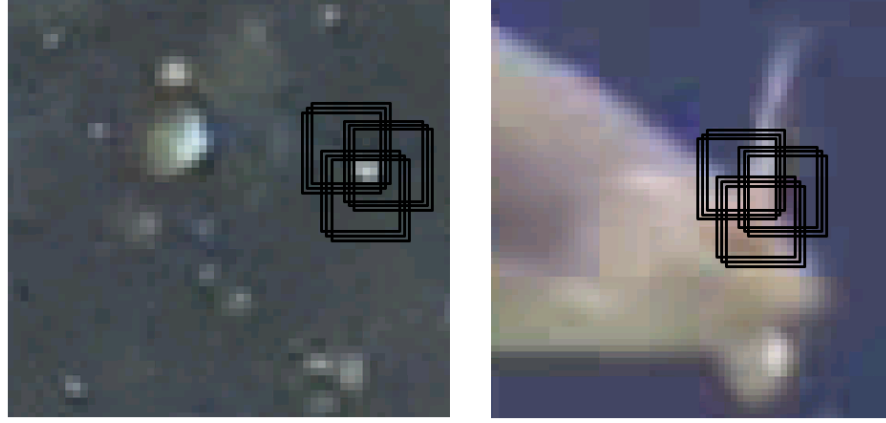


Figure 6.6: Marine snow detection, shows overlapping patches for marine snow (left) versus a true edge (right).

the patch size). Therefore, each pixel of the image could have been filtered in different patches accordingly, which results in having several filtered versions for a single pixel. The final decision about each pixel is then made by using the voting algorithm. If the majority of the filtered versions corresponding to each pixel in the original image indicate that it contains marine snow then this pixel is marked as being corrupted by marine snow and should be filtered out. Otherwise, if the voting algorithm decided that the pixel is not noisy but an object edge, then the pixel value remains as the original value. The ambiguity in such a situation is demonstrated in Figure 6.6.

To estimate the best pixel value, the one which can represent the scene behind marine snow the best, the pixel value in the filtered versions of the current patch, all corresponding to the same location, are rank-ordered and the median value of this list is used as the final approximate. Here it is called Median² filtering, where the final pixel value is the median of median values. Figure 6.7 illustrates the Median² with an example.

Lastly, the algorithm should be adjusted to take care of particles with different sizes. For this, the Gaussian pyramid of the image is obtained in five layers which is a reasonable number for detecting marine particles with maximum size 25×25 pixels. Thus, detection and removal are applied on all these layers. Using this pattern, the smaller particles are eliminated in the first layers of the pyramid while the larger particles are taken care of in the last layers. Furthermore, the small particles that are very close to each other, which could not otherwise be removed in the first layers, are detected together as a large one in the last layers.

6.4.3 Simulation Results and Discussions

The proposed method is applied on several UW images taken in both sea and pool in the presence of marine snow. In some images, such as the one which is taken at Ozeaneum Stralsund (Figure (6.8)), no direct artificial light is used, though those which are taken at the sea (Figure (6.9) and (6.10)), due to depth, used an artificial light. In these cases, the focus is on removing marine snow

6.4 Single Image Marine Snow Removal via Median² Filtering

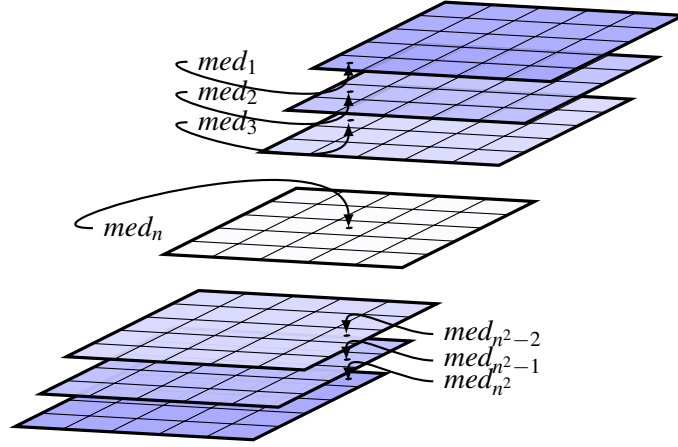


Figure 6.7: The overlapping patches containing a specific pixel are shown, where the pixel identified with a black dot is present in all the patches and detected as marine snow. The value of such a pixel is replaced by median value of each individual patch (med_1, \dots, med_n). The final result, the value of the marked pixel in the output image, is the median of all these median values.

itself rather than strong light reflections.

In addition to the qualitative illustration, the proposed algorithm is compared to several existing approaches. Beside one approach, which directly addresses marine snow elimination [Ban+14], two variations of Median filtering [WZ99; SE07], together with the basic Median filtering, are employed to evaluate the performance of the proposed method. The reason for such a selection is the similarity in the methodology and/or the problem.

Figure (6.8(a)) illustrates a relatively simple example of an UW scene with marine snow and the result after applying the proposed algorithm (6.8(f)). The main focus is to remove particles while preserving the edges of the objects as much as possible. As it can be seen in the result of the proposed method, that marine snow is removed adequately while even the small image features (edges of the fishes) are preserved. It outperforms the three different versions of Median filter, basic Median filter, progressive Median filter [WZ99] and decision-based Median filter [SE07].

This is when the basic Median filtering with window size 5×5 , Figure 6.8(b), eliminate some of the particles but at the same time, it smoothed the image too much which destroyed all the details. Simply decreasing the window size, 3×3 , does not help since it is too small, which has almost no effect in removing marine snow with minimum size 3×3 . The progressive Median filtering (Figure 6.8(c)) removed marine particles partially but clearly could not distinguish between marine snow and object edges thus, the edges of the objects (fishes) were destroyed. Lastly, the decision based Median filtering (Figure 6.8(d)) has completely failed to remove marine snow. This is due to its different criterion in detecting noise which does not match marine snow properties.

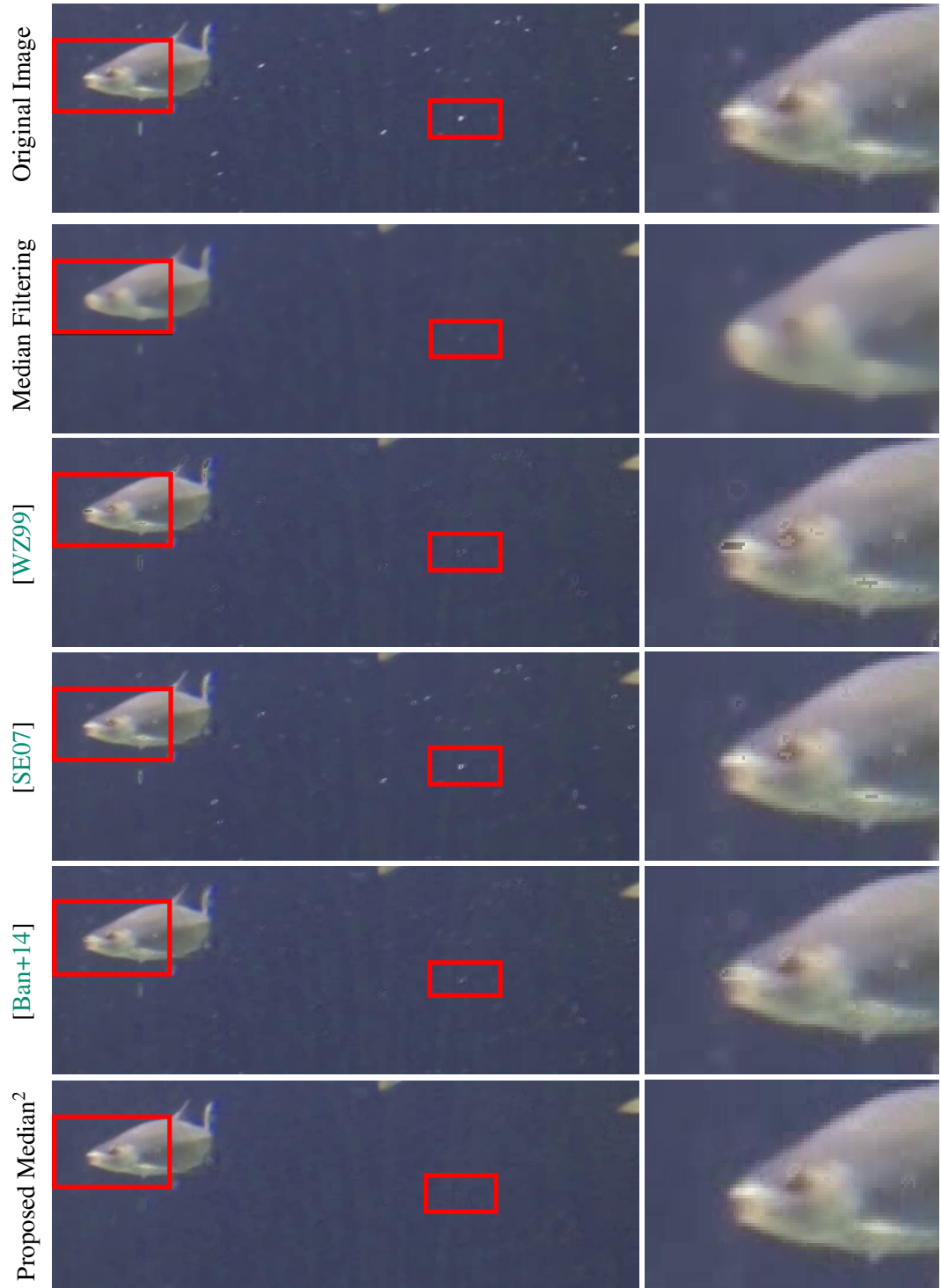


Figure 6.8: The visual comparison of the proposed method with the original image, basic Median filtering, [WZ99], [SE07], and [Ban+14]. Areas assigned with rectangles are illustrated to show the shortcomings of the literature.

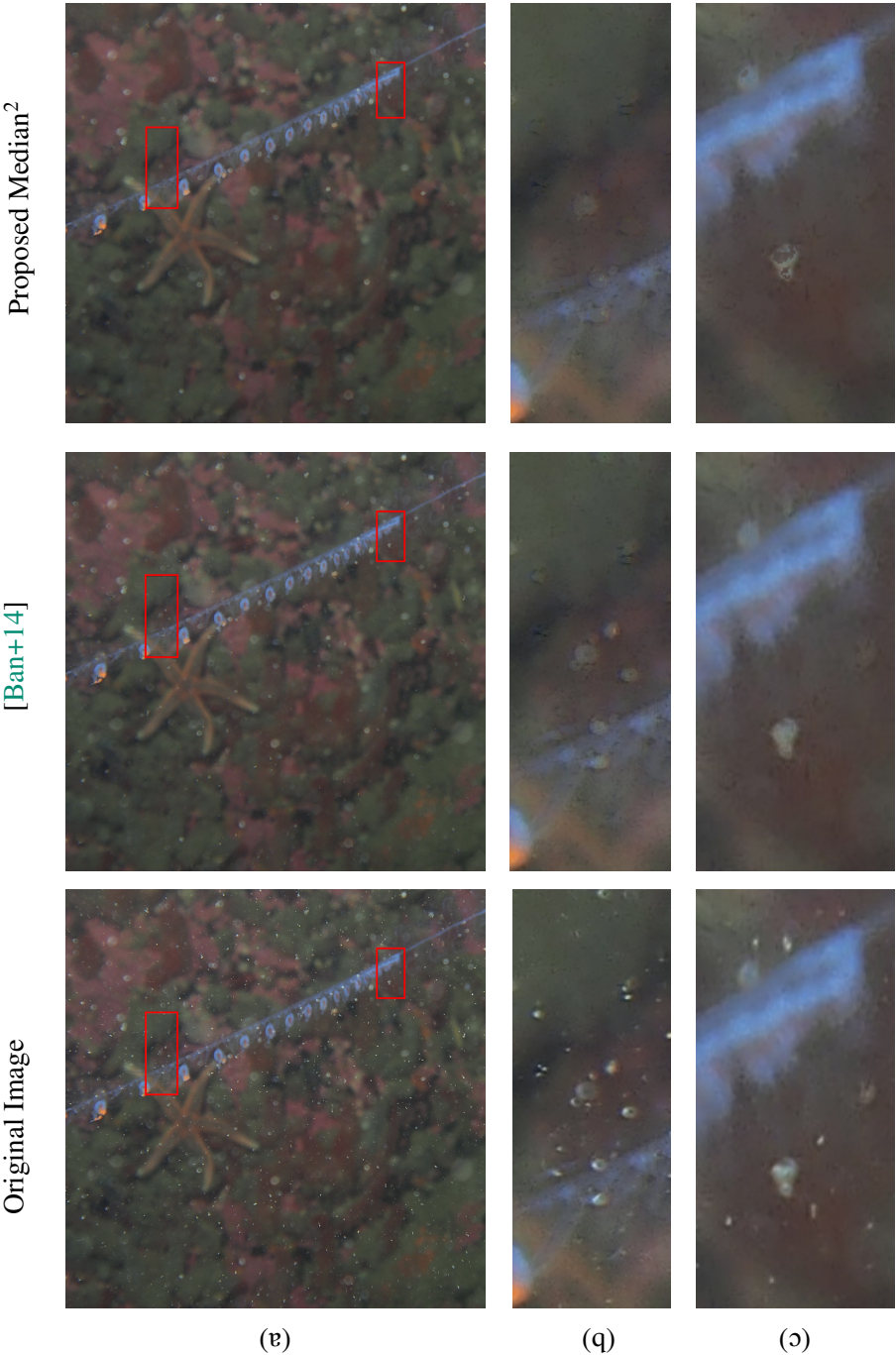


Figure 6.9: Underwater scene with marine snow (courtesy of Eik Deistung). From left to right: the original image, the result after applying [Ban+14] and the proposed Median² approach. (a) illustrates the full image. (b) and (c) show zoom-in presentations of each image. See details in the text.

At first glance, [Ban+14] (Figure 6.8(e)) shows competitive results but as more attention is given to the image, it can be seen that the proposed algorithm has saved the edges of the objects where [Ban+14] has falsely filtered them (e.g. red rectangle). In both algorithms the larger size particles are not filtered completely, yet the proposed algorithm could detect more particles than the approach in [Ban+14].

Figure (6.9) shows the results of the proposed method and [Ban+14] for a more challenging case, where marine snow is more dominant and contains some reflections. This leads to more difficulties where reflections overlap with marine snow and prevent the method from detecting and removing them completely. Despite large reflections which are still present, the proposed algorithm was able to remove marine snow while very small details of marine creature are retained. The reason for such a shortcoming is that the size and intensity of marine snow reflections are different than the marine snow itself and therefore are not considered at detection stage for removal.

Lastly, a particularly difficult case (Figure 6.10), highly corrupted with marine snow where the light reflection on the particles introduced more difficulties, is examined. The proposed approach could eliminate marine snow effectively. This is when the result of [Ban+14] could not successfully remove marine snow (especially the larger size particles). The zoom-in presentations show the promising result obtained using the proposed algorithm, although, due to using an artificial light source, the reflection of marine snow is dominant enough to disturb the visibility of the image even more. Thus, in the full version of the result, one can observe that there are still some reflections of marine snow particles which lead to an unpleasant look of the image.

In addition to qualitative evaluation, the proposed algorithm is assessed using PSNR and MSE. This requires a reference image (ground-truth data); however, it is usually not available in the UW situation. In such a situation, it is a common practice to apply the algorithm on simulated data for quantitative evaluation where the qualitative performance of the algorithm can be evaluated on both real and simulated data.

To this end, one needs to model marine snow properties which remains as a subject under investigation [SBR11]. A common approach used in the literature is to simulate marine snow by generating salt & pepper noise on the images. However, as discussed previously, this model does not take into account various physical parameters of marine snow such as the size and shape of the particles, the defocus effect or the effect of water absorption and scattering on the signal backscattered by the particles.

Boffety and Galland [BG12] proposed a method to model this phenomenon where the effect of water, and backscattering signal are considered. This is done by assuming that these particles behave like white *lambertian* scatters. The reflection coefficient of particles is presumed to be a Gaussian function. Lastly, they positioned such simulated particles in the image according to a salt & pepper distribution and computed the backscattering component for each volume of water. This model takes into account some physical properties of the particles as well as water absorption and scattering effects; however, it still does not cover the differing geometries of marine snow, thus,

6.4 Single Image Marine Snow Removal via Median² Filtering

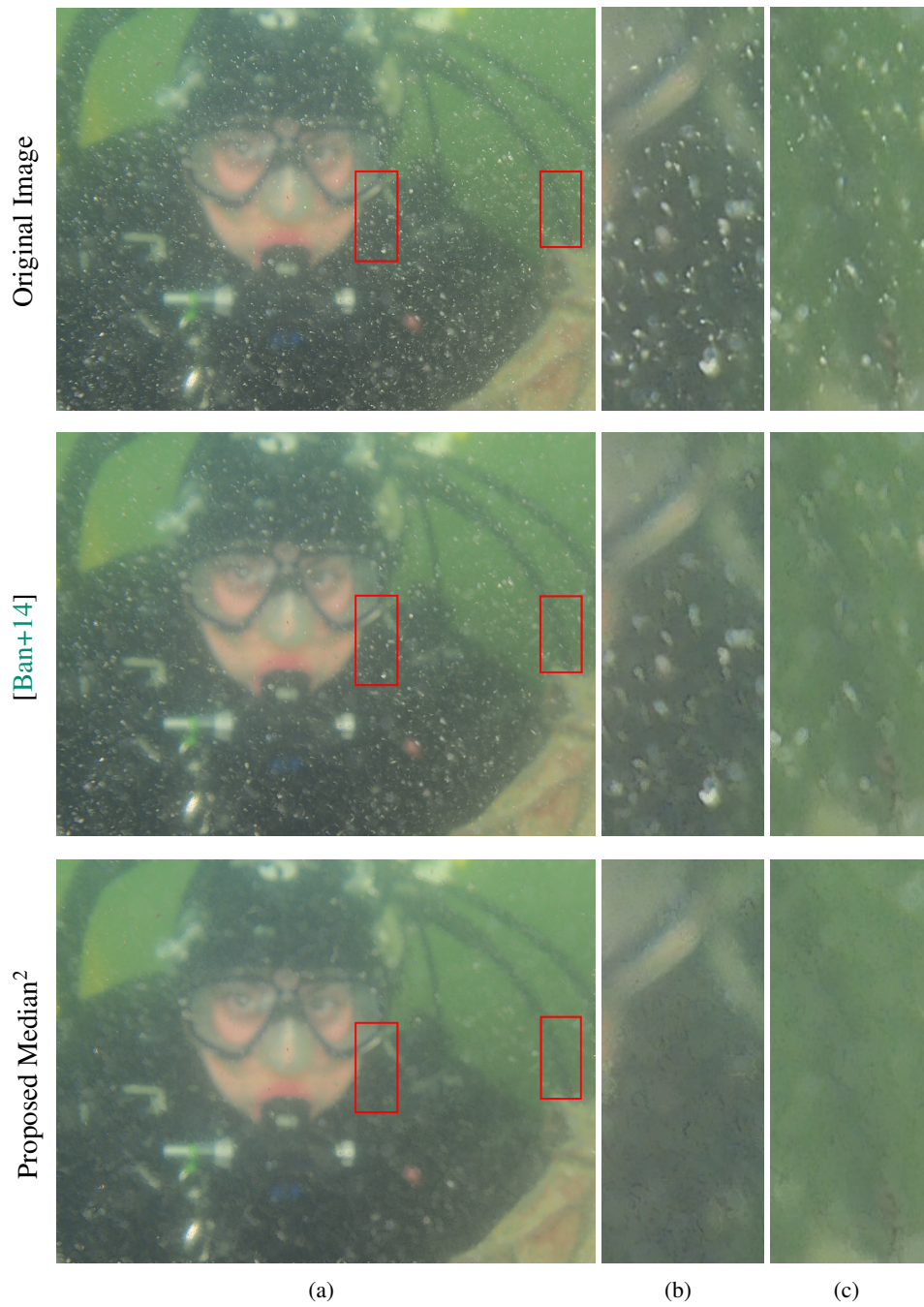


Figure 6.10: Underwater scene with marine snow (courtesy of Eik Deistung). (a) illustrates the full image. (b) and (c) show zoom-in presentations of each image. See details in the text.

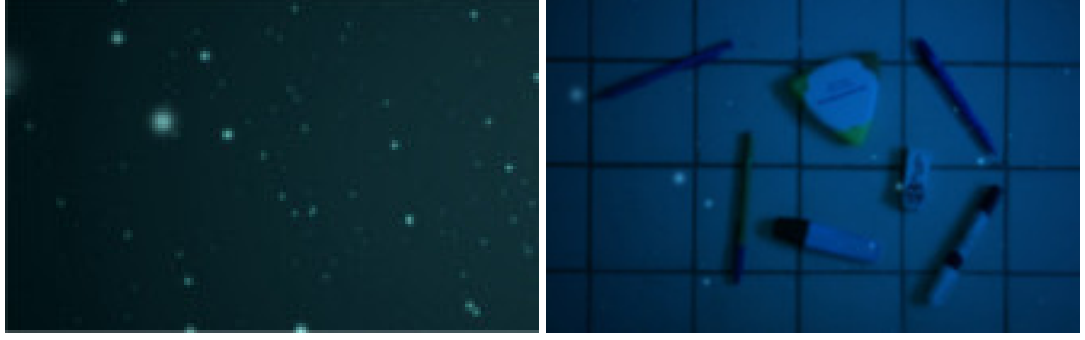


Figure 6.11: Examples of simulated images corrupted with marine snow based on the proposed model in [BG12]. Images are courtesy of [BG12].

it gives an artificial look to the image. Examples of such simulated data can be seen in Figure 6.11.

Here, using the same strategy but considering different geometries for particles, marine snow is simulated. Moreover, a second strategy is also used to provide synthetic data. In contrast to the common approaches where the degradation is simulated on an existing ground-truth image, here, a degraded image, containing marine snow, is used to first conduct the ground-truth image. Next, the degradation effect is simulated on this reference image.

To do so, for each test image corrupted by marine snow first, the particles are extracted precisely. This is done by human experts where particles are detected and segmented out manually, resulting in a pixel-wise accuracy. Once all particles are extracted, the scene corresponding to missing pixels is restored using the information provided by the neighborhoods. This image is then called the "ground-truth image". The next step is to simulate marine snow in this ground-truth image. This is accomplished simply by randomly positioning extracted marine snow from the original image in the ground-truth image. This simulated image, together with the ground-truth image, is then used to evaluate the performance of the proposed method. This way, one can obtain simulated data with a very natural look where marine snow has the most accurate model and is highly correlated to the real scene. Examples of such simulated data can be seen in Figures 6.15 and 6.16, termed Stralsund and Geomar data respectively.

Table 6.1 illustrates the comparison of the results on the Stralsund and Geomar data using basic Median filtering, [WZ99], [SE07], [Ban+14], and the proposed method respectively. The evaluation is done using PSNR and MSE. In view of this table, one can observe that the quantitative results are inline with the outcomes of visual analysis where the proposed Median² algorithm outperforms the existing algorithms having the highest PSNR and lowest MSE values.

Both visual and numeric evaluations suggest that the proposed algorithm could successfully remove marine snow from single images where it preserves edges as much as possible. Compared to the method proposed in [Ban+14], the Median² approach achieves better results, especially for images with higher corruption due to the presence of marine snow (the more difficult cases).

6.4 Single Image Marine Snow Removal via Median² Filtering

PSNR/MSE Values		
Approach	Stralsund Data	Geomar Data
Basic Median Filtering	32.5911 / 31.6520	25.3145 / 56.2471
[WZ99]	33.9619 / 29.0043	24.9858 / 60.9021
[SE07]	28.5264 / 44.9482	24.7343 / 62.592
[Ban+14]	43.6504 / 1.6750	37.4235 / 3.4235
Proposed Median ² Filtering	47.4818 / 1.0776	43.7806 / 1.6501

Table 6.1: Evaluation using PSNR and MSE values corresponding to the simulated images from Stralsund and Geomar data, provided in Figures 6.15 and 6.16 respectively.

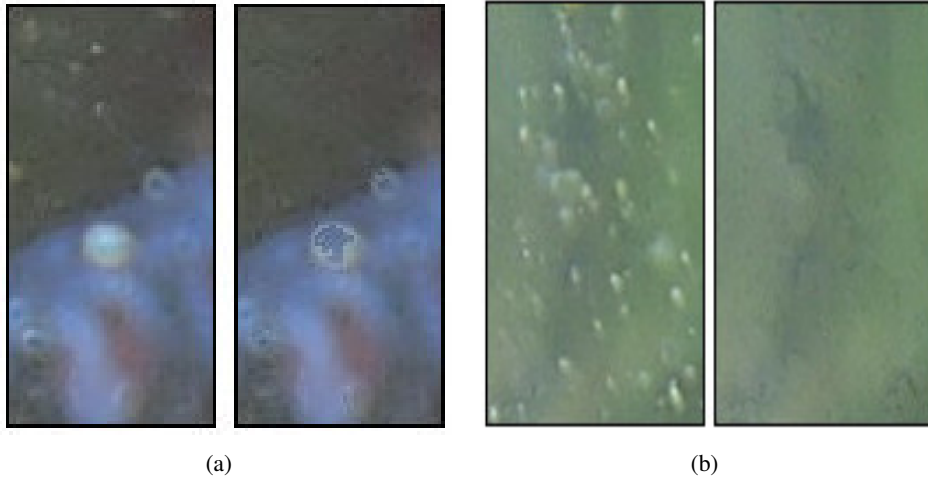


Figure 6.12: Examples of the problems that may occur using Median² approach. (a) when detection is not precise, and (b) when marine snow is dense so not much information is available. The images are a zoom-in presentation of the original image and the result of the proposed method in Figure 6.9 and 6.10 respectively.

The efficiency of the algorithm depends on several factors such as image pixel resolution, the number of marine snow particles and the number of layers in the image pyramid. Since the particular goal is to achieve the best image quality, the algorithm is optimized for quality though not for speed. Furthermore, the images are employed with the highest pixel resolution and the maximum number of necessary image pyramid layers. Empirically, 5 layers of an image pyramid has shown to be optimal for full HD images (1920×1080). Depending on the degree of corruption (number of particles), the algorithm needs between 4 to 20 seconds for full HD images.

Notwithstanding the promising outcome, the proposed algorithm may not give the best result in some scenarios such as:

- When the detection is not perfect which usually happens when particles have larger sizes. Marine snow is not detected perfectly; thus, the remaining pixel values in the patch are partially corrupted by marine snow. In such a situation, filtering may lead to a wrong approximation and thus marine snow is not removed completely. Examples of such a situation can be seen in Figure 6.9 (c) and 6.12 (a).
- When marine snow is dense and therefore, not much information in the neighborhood remains in order to accurately estimate the scene behind it. Figure 6.10 (c) and 6.12 (b) show examples of such a situation.
- It is well-known that filtering may cause blurriness. Although, here Median filtering is used, yet, the approach may blur the image especially for highly corrupted images.

In image processing algorithms where there is an uncertainty or noise which can affect the performance of the algorithms, prior models are commonly used. The prior models of image structure are used in many machine vision problems such as stereo, optical flow, denoising and super-resolution to name a few. These priors can be provided by means of a depth map or flow fields given together with the data. However, such priors are not always available, thus, learning algorithms such as the Markov Random Field (MRF) or sparse representation have attracted a lot of attention. Here, in case of marine snow, due to the fact that there is no available information behind rather large structures (particles with sizes approximately between 3×3 and 25×25), modeling the statistics of the images can help the restoration of the scene. This leads to a more accurate result where the shortcoming of image filtering, listed above, can be overcome. The image priors can be learned using an arbitrary data set of the same class as the input image, however, once the available data is more than single images, one can learn the image priors from the frame sequence. This leads to highly correlated learned priors to a specific scene. This could be beneficial especially for UW data where many factors, i.e. water type, affects the captured data which may make it unique in the sense of image statistics.

6.5 Marine Snow Removal from Video Frames with a Static Camera

Here, a different perspective at removing marine snow is considered. Having videos with a static camera, where some movements such as tilt, pan or zooming are also allowed, the proposed algorithm will detect and eliminate marine snow from the video frames. This is done via an image restoration algorithm where the information provided by the video sequence is used to recover the scene covered by marine snow.

To do so, first, a background modeling algorithm is used to obtain an accurate model of the static components of the video. This model provides the information about the background which is covered by marine snow. For this, a mixture Gaussian model extracting the static components of the scene is introduced. This will be explained in detail in Section 6.5.1.

Concurrently, for each individual frame, a mask defining locations of marine snow is obtained. The approach proposed for single image marine snow removal (explained in Section 6.4) is

employed in order to provide such a mask. This is done as follows: once the corrupted pixels are detected, instead of Median² filtering, the corresponding pixels are marked in a mask. The mask is a matrix the same size as the original frame and contains binary values; one for the locations of marine snow and zero elsewhere.

Next, the background model is used as training data and some priors about the scene are learned. *Field of Experts* [RB05] is employed for this purpose. The detailed explanation of this method is provided in Section 6.5.2.

Lastly, marine snow is eliminated through an inpainting algorithm where the extracted mask and the learned image priors are used to recover the scene covered by the particles.

6.5.1 Background Modeling

Background modeling and subtraction is a common computer vision method for the detection and segmentation of moving objects in videos captured with a static camera. The basic background modeling is based on the average [LH02] or the median [MS95] or histogram analysis over time [Zhe+06].

However, most of the current algorithms use statistical methods for the modeling process. Among others, Gaussian distribution is most commonly used to model the color information of frame pixels in a video sequence and to extract one image which only contains the background. For this purpose, one can use a single Gaussian [Wre+97]; yet, such a model cannot handle dynamic background, i.e. when there is water rippling. To overcome this problem, the Mixture of Gaussian (MOG) [SG99; Wan+14] has been proposed. Using MOG leads to achieving high accuracy results; however, it also has a downside. Because it is not a trivial task to unify the different Gaussian distributions. Thus, to keep the balance between accuracy and complexity, a Gaussian Switch Model (GSM) has been proposed which makes use of the advantages of both approaches.

The idea behind this algorithm comes from the shortcoming of a single Gaussian approach which includes the information from foreground objects in the background model and then corrupts it. This happens especially when there is a constant presence of many foreground objects. This can be solved by applying a *partial update*, this means that instead of updating the whole model, only the pixels that are classified as background are updated. Ideally, now only background information is included in the model which should lead to a more robust and precise model. The principles are as follows: at each iteration before updating the background model, a segmentation of the current frame is computed by background subtraction. The segmentation is then used to exclude foreground objects and the algorithm only updates the background pixels. In general, this improves the segmentation and stabilizes the model, but since the model is used to improve its updating process itself, what could be deemed as self-fulfilling prophecy can occur.

An example of this is the presence of a foreground object in the initialization. This foreground object is a part of the model in the beginning and should slowly be overwritten with

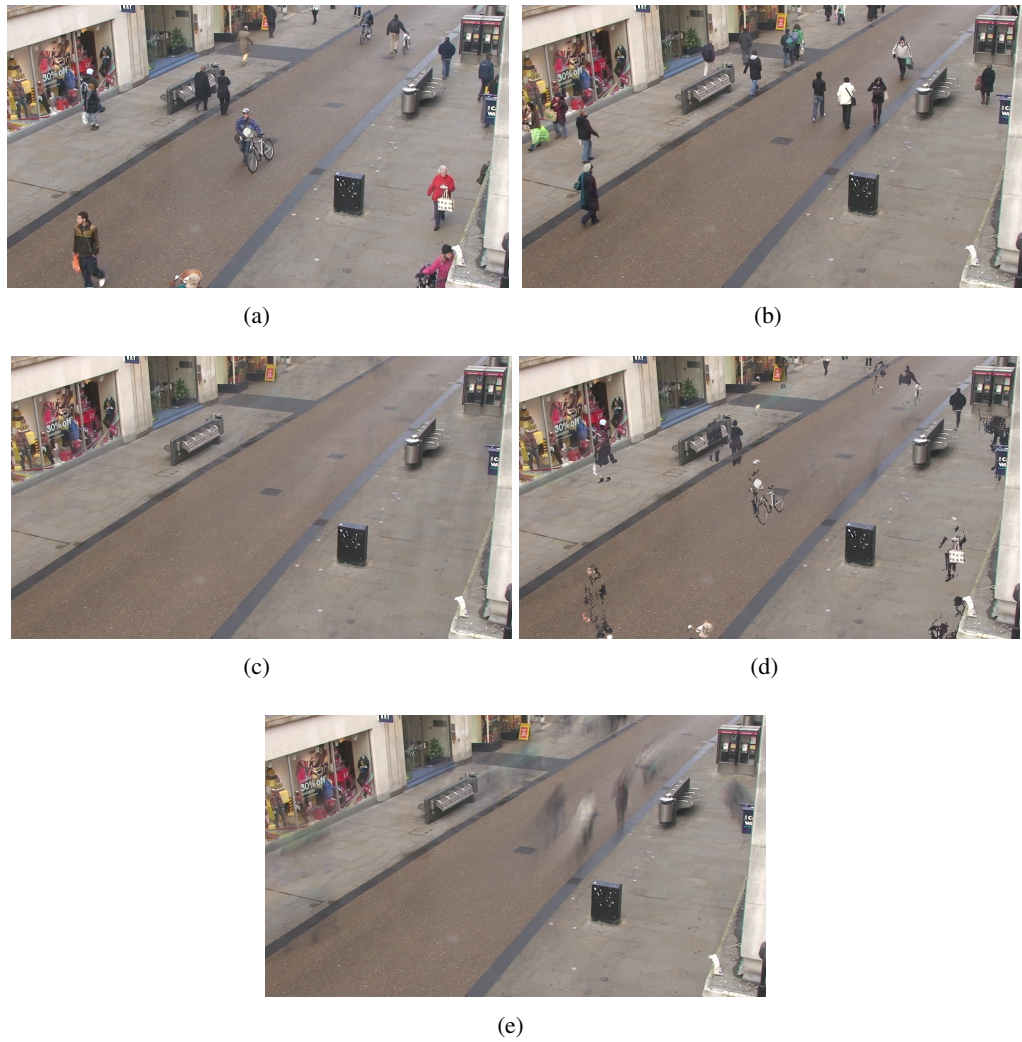


Figure 6.13: Comparison of different update schemes for the background modelling. In the top row are the first (a), and 2000th (b) frames of the Town Center Video [BR11]. Second and the third rows correspond to the background models for the 2000th frame created with different updating mechanisms: (c) GSM, (d) partial updating and (e) a complete update for every frame.

the background information during the updating process. However, when partial updating is applied, this usually does not happen because the actual background in that area will be marked as foreground; therefore, it does not get included in the model. Figure 6.13(d) provides such an example where the background model obtained using partial update still contains foreground objects.

To still get the benefits from the partial updating without facing these problems, the proposed GSM uses two Gaussians to model the background. The first Gaussian is partially updated and

is taken as the background model where the second Gaussian is fully updated with every frame. The errors of the partial updated model can be discovered by a comparison between these two Gaussians, since they always show the same characteristics:

- The means of the two Gaussians slowly diverge from each other as the Gaussian with the full update adapts to the new background and the other remains constant.
- For many successive frames a foreground object is detected in the same position.

If these characteristics are true for a specific pixel, the partially updated Gaussian for that pixel is overwritten with the values of the full updated Gaussian as it does not reflect the true background any more. Furthermore, the switching to the fully updated Gaussian only happens when the variance of this Gaussian is small because this indicates that there has not been much noise or many foreground objects recently and therefore this Gaussian can be a reliable background model. The represent version of the background model can then be simply extracted by taking the mean of the partially updated model for each pixel and color channel.

An example of the background modeling with the GSM compared to the partial and full update approaches can be seen in Figure 6.13 (the mean values of the Gaussians are displayed) where it is compared to the full and partial updating schemes on a video with many foreground objects. The parameters of the modeling are the same for all three methods and it can be seen that the complete update created a model which is very corrupted with the current foreground objects of the scene. The partial update eliminates this problem but many objects from the first frame can still be seen in the model as they are never eliminated. The GSM can combine the advantages of both methods and creates an almost uncorrupted background model.

In addition, the proposed method needs to be adapted to some camera movements. This is crucial especially for UW scenarios where the robot (i.e. ROV) is floating and may move up and down due to waves; as a result, the camera is not completely static as it is assumed to be. Such movements can also occur when i.e. a fish hits the camera; thus, there will be a sudden and short movement. Another useful situation is zooming where the diver or ROV is not able to get close enough to the object of interest; thus, the diver or ROV stops at distance but zooms for better perception. Lastly, this adjustment can be used for a slowly moving robot which can be treated as a continues pan/tilt movement.

Therefore, pan, tilt and zoom are considered as the most probable movements, thus, the background modeling algorithm is adapted accordingly in a way that the background model is not influenced. For this, an extension of BM is proposed. First, the camera movement is detected and second, computing the affine transformation of the background model to the current scene, the model is adjusted accordingly.

The algorithm detects the movement by comparing the frame t with $t - 2$. The distance between the two frames can vary based on the speed of the camera motion, pan or tilt, and the frame rate of the video (generally, the minimum distance of two frames is a fair choice since usually, such movements last longer than two frames). If the comparison reveals a movement, then the background model obtained in the last frame is modified accordingly to match the scene in the

current frame. This is done by applying different affine transformations, the ones corresponding to pan, tilt and zooming, on the available background model. The result is then compared to the current frame (t) and the best match is taken as the true transformation. Once the true transformation is attained, the background model is transformed accordingly. The unknown area in the updated background model is filled with the corresponding pixel values from the current frame (t) and finally, this area is gradually updated with the background.

Thus, the final background model matches the current frame in case of any aforementioned movements. Applying this on UW videos, it has been observed that, as long as the scene contains some features which could be used to detect the movement, the algorithm could fine-tune the background model to the current frame.

Having an accurate background model, the next step is to learn statistics of the scene which can ease the restoration task.

6.5.2 Statistical Priors Modeling using the Background Model

Considering marine snow as a source of noise, although, it is an object in the scene but nevertheless unwanted, introduces many difficulties into elimination of particles. This is equivalent to restoring the scene behind small objects where there is no information available. Thus, learning image statistics of not only image patches but also their relation (leading to a rich prior model for the whole image) is the most appropriate attempt. This is due to the fact that, in case of heavy marine snow, the uncertainty is not only in local patches but their connectivity with each other.

There are different approaches attempting to model image priors. In many works such as [HS08; Dab+08; Mai+09b] priors are learned over small image patches. This makes computational tasks, e.g. learning, inference and likelihood estimation much easier than working with the whole image. Examples of such approaches which address modeling complex image structure within patches, are sparse coding algorithms.

Sparse coding represents structural properties of image patches in terms of linear combinations of filter responses:

$$\underset{\vec{\alpha}}{\operatorname{argmin}} \|\vec{x} - D\vec{\alpha}\|_2^2 \quad s.t. \quad \|\vec{\alpha}\|_0 < T. \quad (6.9)$$

Where $\vec{x} \in \mathbb{R}^N$ stands for vectorized image patch, $\|\cdot\|_0$ counts number of nonzeros which is controlled by threshold T , $D \in \mathbb{R}^{N \times K}$ denotes the filter bases and $\vec{\alpha} \in \mathbb{R}^K$ is sparse coefficient vector.

Despite providing rich priors modeling local image structures, since these methods focus on image patches, they therefore do not provide a direct way of modeling the statistics of whole images.

On the other hand, image priors can be obtained via Markov Random Fields (MRF). MRF is a graph $G = (V, E)$.

- $V = \{1, 2, \dots, N\}$ is the set of *nodes*, each of which is associated with a random variable, u_j , for $j = 1, \dots, N$.
- the neighborhood of node i , denoted N_i , is the set of nodes to which i is adjacent; i.e. $j \in N_i$ if and only if $(i, j) \in E$. E stands for the edges connecting nodes.
- The MRV satisfies

$$p(u_i | \{u_j\}_{j \in V \setminus i}) = p(u_i | \{u_j\}_{j \in N_i}). \quad (6.10)$$

Definition 6.1 A graph $G = (V, E)$ is a finite collection of nodes V (here pixels) and set of edges E .

The key to MRF is that, through local connections, information can propagate a long way through the graph. However, MRF priors crudely capture the statistics of the natural images where they exploit handcrafted clique potentials and small neighborhood systems [GG84]. Consequently, it does not provide rich priors as patch-based algorithms can do.

Roth and Black [RB05] proposed a framework to learn image priors which combines the advantages of both algorithms, sparse modeling and MRF, and learns rich and generic prior models of natural images (or any class of images). This so-called *Field-of-Experts* (FoE) applies the idea of *Product-of-Experts* (PoE) approach [Hin99] and extends it to be able to learn generic priors for images of arbitrary sizes.

Product-of-Experts

A very efficient way to model high-dimensional data is to combine the multiple probabilistic models of the same data which satisfy different low-dimensional constraints.

Product of Experts offers an alternative to combining many individual expert models, where each expert works on a low-dimensional subspace that is relatively easy to model. This is done by multiplying experts together and then renormalizing, and as a result, producing much sharper distributions than the individual expert models. Examples of selected expert models are a mixture of Gaussians, a mixture of a Gaussian and a uniform and many other popular models such as linear dynamic systems. PoE combines n individual expert models as follows:

$$p(\vec{x} | \Theta) = \frac{1}{Z(\Theta)} \prod_{i=1}^n p_i(\vec{x} | \theta_i) \quad \Theta = \{\theta_1, \dots, \theta_n\}, \quad (6.11)$$

where $p_i(\vec{x} | \theta_i)$ is the probability of data vector \vec{x} under model i , and θ_i is all the parameters of an individual model i . $Z(\Theta)$ stands for normalizing function.

Although PoE provides an elegant and powerful way of learning image priors on small image patches, the results do not generalize directly to give a prior model for the whole image. One cannot generalize the model by simply increasing the patch size since:

- the number of parameters to learn will be too large.

- the model works for a specific image size and does not generalize to other sizes.
- the model is translation variant.

Field-of-Experts

Roth and Black [RB05] proposed an extension of PoE which overcomes the aforementioned difficulties. They extended traditional Markov random field models by learning potential functions over trained filter bases. To this end, they employed a set of external example images and developed a *parametric representation* which uses the examples for training rather than using them directly for representation. Sparse representation is used to define experts which are image patches projected onto linear components of the dictionary. This is equivalent to filtering a patch with a *linear filter* described by dictionary atom (d_i).

To be more precise, consider the pixels in an image X to be presented by nodes V in a graph $G = (V, E)$, where E stands for the edges connecting nodes. A rectangle region of $m \times m$ neighborhood connecting all nodes is defined, where every such neighborhood is centered on a node (pixel) defining a maximal clique $X_{(k)}$ in the graph (for $k = 1, \dots, K$).

Definition 6.2 *clique is a complete subgraph of G . A maximal clique is the one with a maximal number of nodes; one cannot add any other node while still retaining complete connectedness.*

They made two main assumptions. First, the MRF is homogeneous, meaning the potential function is the same for all cliques, in other words, $V_k(X_{(k)}) = V(X_{(k)})$. This leads to translation invariance of the MRF model. The second assumption is made so that the maximal cliques in the MRF potentials are square patches of a fixed size. This does not affect the method generality since other, non-square, neighborhoods could also be used [GR92].

The novelty of this approach is where they extended MRF beyond **FRAME** and learned the potential function from training images instead of defining it by hand. This is done by representing the MRF potentials as PoE where sparse representation is used to define experts ϕ_i :

$$p(x) = \frac{1}{Z(\Theta)} \prod_{i=1}^n \phi_i(d_i^T X_{(k)}, \alpha_i) \quad \Theta = \{\theta_1, \dots, \theta_n\}, \quad (6.12)$$

and

$$\phi_i(d_i^T X_{(k)}, \alpha_i) = (1 + \frac{1}{2}(d_i^T X_{(k)})^2)^{-\alpha_i}, \quad (6.13)$$

here d_i is a linear filter (dictionary atoms), and n stands for the number of experts (ϕ_i). As a result, the probability density of a full image is obtained as follows:

$$p(X) = \frac{1}{Z(\Theta)} \prod_k \prod_{i=1}^n \phi_i(d_i^T X_{(k)}, \alpha_i), \quad (6.14)$$

The parameters α_i and linear filters d_i are learned from the set of training images. The main difference with respect to the PoE is that FoE takes the product over all neighborhoods k .

As a result, the model works for different image sizes, enjoys a translation invariant property which is desirable for generic image priors and has few parameters which need to be learned. The key advantage of the FoE model is that it gives a way to build richer prior models that combine more filters over larger neighborhoods while also doing this in a principled way.

6.5.3 Inpainting

Once the FoE priors are available, the inpainting algorithm [RB05] propagates information using only these priors and refill the pixels defined in mask M iteratively by introducing an iteration index t and an update rate η as follows:

$$X^{(t+1)} = X^{(t)} + \eta M \sum_{i=1}^N d_i^- * \Psi_i(d_i * X^{(t)}), \quad (6.15)$$

where

$$\Psi_i(y) = \frac{d}{dy} \log \phi_i(y, \alpha_i) \quad (6.16)$$

let X to be the inpainted image, d_i^- denotes the filter obtained by mirroring d_i around its center pixel [ZM97], and $*$ stands for convolution.

6.5.4 Marine Snow Removal

To finally eliminate marine snow from frames of videos, equivalently, to restore the scene covered by the particles, the following steps are proposed:

- **Background Modeling:** using the GSM algorithm [RG15] a background model of the scene is obtained. This reference image is an approximation of the scene covered by the marine snow.
- **Mask Extraction:** in the meantime, employing the proposed single image marine snow removal via Median² approach, a mask indicating the positions of particles has been acquired. In that sense, steps reminiscent of the ones in the Median² are encountered. After the coarse detection is completed, the introduced voting algorithm makes the final decision about the corrupted pixels (the ones containing marine snow). In this stage, instead of applying Median² on the corrupted pixels in the original image, a mask is obtained which defines their locations.
- **Learning Prior Models:** employing the FoE approach [RB05] together with the background model (as the training data), some priors about the statistics of the scene are learned.
- **Inpainting:** having the priors together with the extracted mask, an inpainting algorithm is used to restore each frame without marine snow particles.

Employing the background model as the training data for learning prior models takes into account the temporal information of the video, thus restoration can be applied on each frame separately.

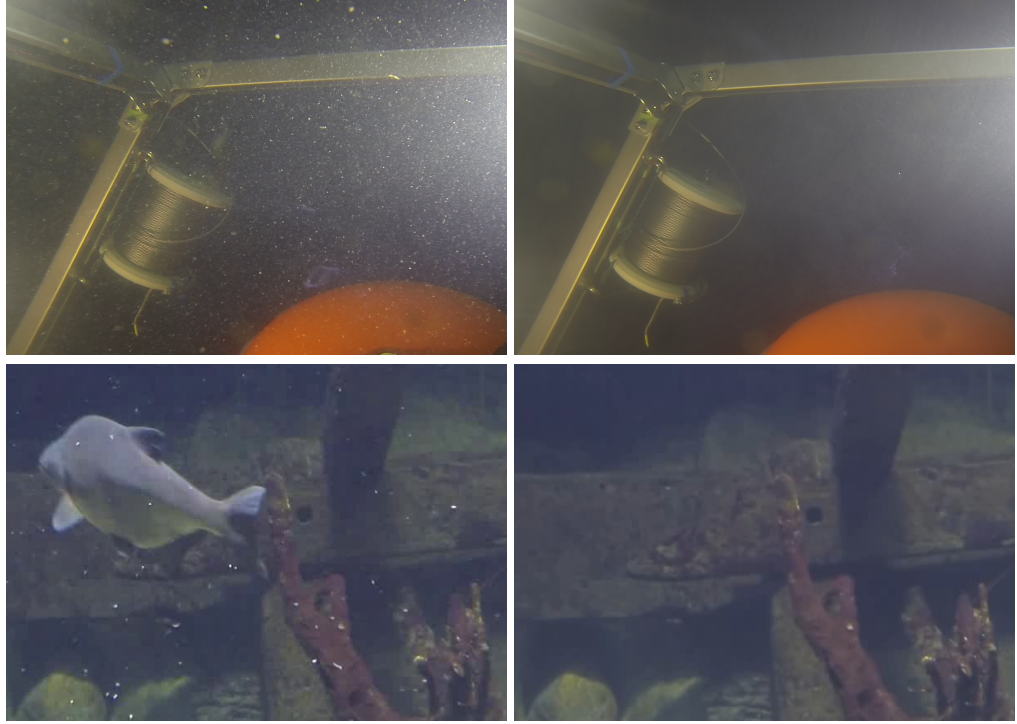


Figure 6.14: Examples of the background modelling. First column illustrates an example frame of the original video, and second column shows the background model corresponding to each scene obtained using GSM.

6.5.5 Simulation Results and Discussions

The proposed algorithm is performed on different UW videos with marine snow. The videos are taken in both pool and sea. As examples, the scene in Figure 6.15 is taken at Ozeaneum Stralsund, and the video corresponding to Figure 6.16 is taken in the Black Sea (courtesy of GEOMAR).

Generally, a reasonable number of frames are used to train a background model of each video (training set). Usually, the first 500 frames are the choice. Nevertheless, this may vary depending on the length of videos. If the length of a video is not long enough to support such a decision, then a background model can be obtained using fewer frames. However, to assure the highest accuracy of the acquired model, one can extend the number of training frames by duplicating the available data. The Geomar data presented in Figure 6.16 is such an example where the video lasts only 200 frames.

Besides qualitative evaluation of the algorithm, PSNR and MSE are used to provide quantitative measures of the proposed algorithm performance. The same as before, a ground-truth image and a synthetic frame are provided for each video. This is done via manually extracting marine snow from the original frame and next placing the particles in the ground-truth frame. Examples of simulated data can be seen in Figure 6.15 and 6.16. The results are compared to the method in

PSNR Values		
Approach	Stralsund Data	Geomar Data
[Ban+14]	43.6504	37.4235
Proposed Median ² Filtering	47.4818	43.7806
Proposed Restoration algorithm without BG Model	48.6403	47.1723
Proposed Restoration algorithm with BG Model	48.8275	47.6833
MSE Values		
[Ban+14]	1.6750	3.4306
Proposed Median ² Filtering	1.0776	1.6501
Proposed Restoration algorithm without BG Model	0.9430	1.1167
Proposed Restoration algorithm with BG Model	0.9229	1.0529

Table 6.2: Quantitative evaluation using PSNR and MSE values.

[Ban+14], the result of single image marine snow removal via Median² approach (Section 6.4), the proposed algorithm without BG model where inpainting is trained using an arbitrary training set of the same class (UW images) instead of using BG model, and finally the proposed algorithm with BG model where inpainting is trained using the background model (using a highly correlated training data to the testing data). The comparison is supplied via PSNR and MSE calculation in Table 6.2.

In examining of Table 6.2, one notices that detection and filtering marine snow (the proposed method in Section 6.4) has already succeeded by about a 4 and 6.3 dB improvement on the results of [Ban+14] for simulated images of Stralsund and Geomar data respectively. The advantageous of using an inpainting algorithm together with the extracted mask has been proved by achieving further improvement of 1.2 and 3.4 dB accordingly. This is when statistical priors of the scene are used instead of directly using neighborhood pixels. This means that, no matter whether a single image or video frame approach is used, the use of prior knowledge about the scene improves the result, as is expected. However, as can be seen, the training data for the inpainting algorithm plays an effective role, where the result of the proposed algorithm differs when the training data changes. When the training data has a high correlation with the input image (with BG model, using the BG model learned over video frames of the very same scene), the algorithm can achieve about 0.2 dB for Stralsund data and 0.5 dB for the Geomar data improvement compared to the situation where a set of arbitrary images of the same class (UW) are used (without BG model). This illustrates the impact of available data, indeed, once the data is more than single images, one can achieve higher accuracy.

Figures 6.15 and 6.16 illustrate the qualitative results corresponding to table 6.2. The improvement over [Ban+14] is clear, where the edges are smoothed and marine snow is not removed completely. The result of the proposed Median² filtering (Section 6.4) has already succeeded in removing marine snow effectively without smoothing the edges too much. However, a closer look demonstrates the shortcoming of this approach. It does not remove marine snow correctly where it is placed on the edges of the objects or where there are no edges but the intensity values

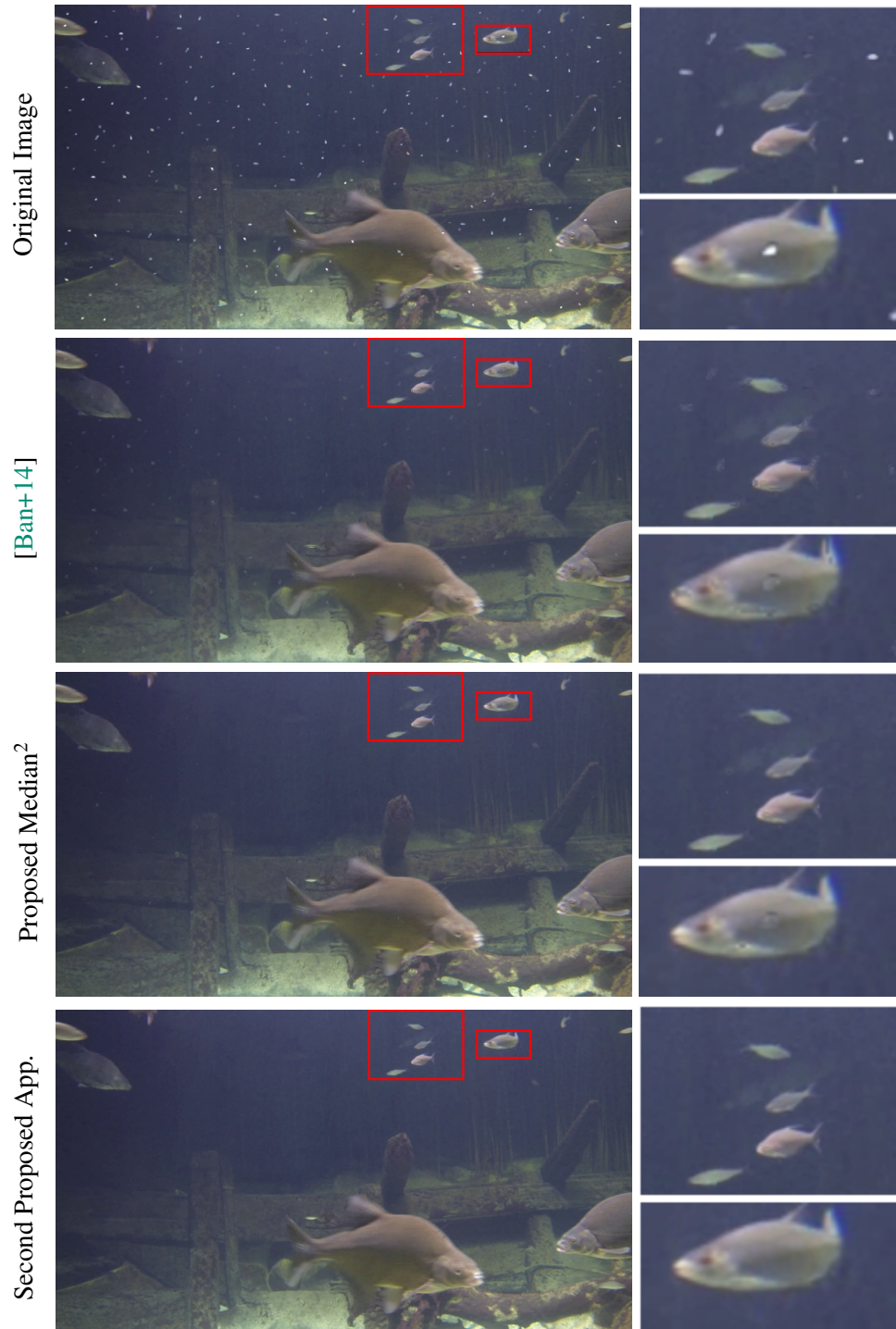


Figure 6.15: Performance of the proposed video marine snow removal for a frame of the Stralsund video. From top to bottom: corrupted simulated image, result of [Ban+14], the proposed Median² filtering and result of the proposed restoration algorithm via inpainting using the BG model. Zoom-in presentations are provided for a better insight.

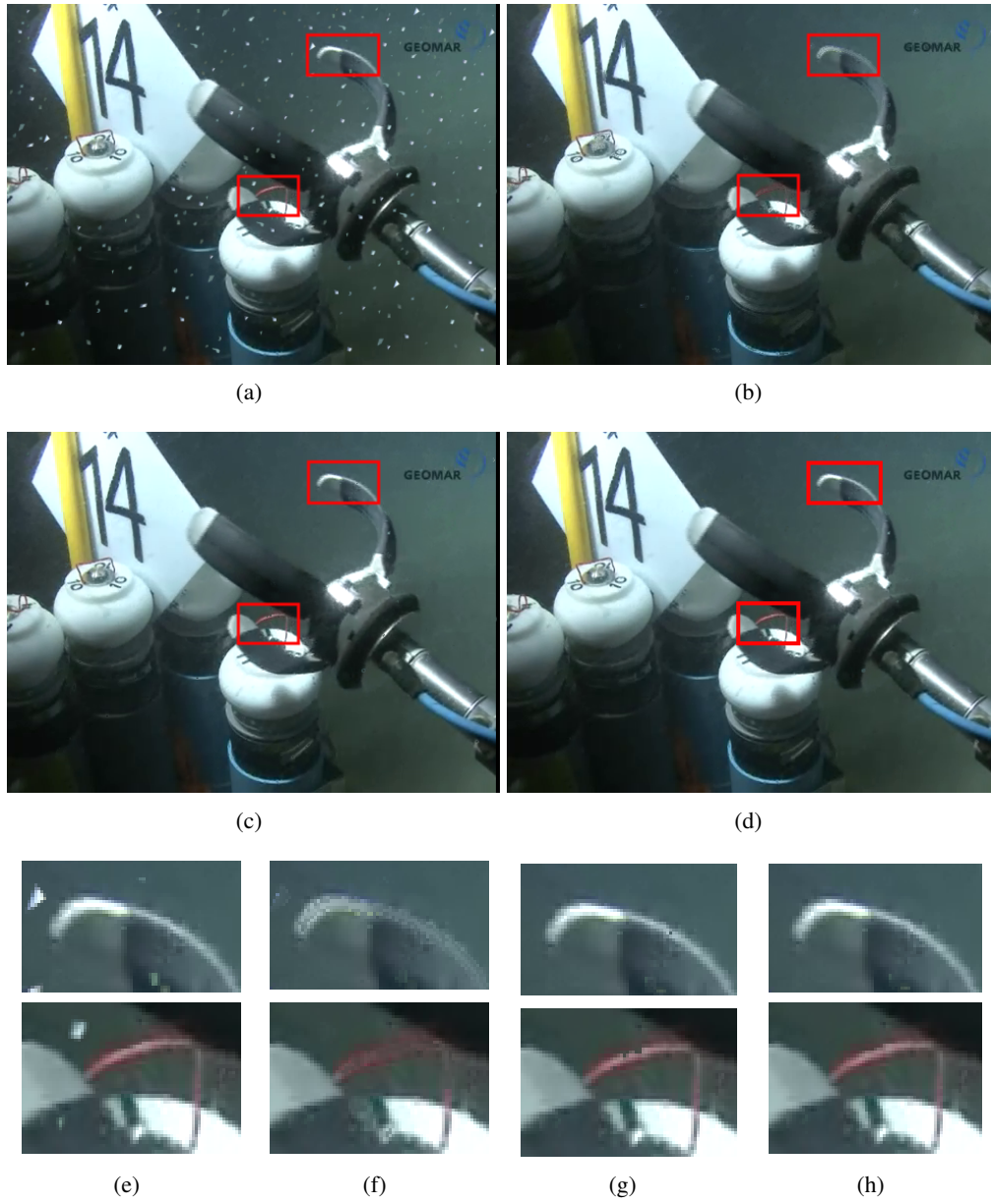


Figure 6.16: Performance of the proposed video marine snow removal for a simulated frame of the Geomar video. (a) corrupted synthetic image, (b) result of [Ban+14], (c) the proposed Median² filtering and (d) final result of the proposed restoration algorithm via inpainting with BG model. Two zoom-in presentations of each image are shown for a better comparison (e-h) respectively. Full video of this scene is courtesy of JAGO-Team, GEOMAR Kiel, Germany.

changes (e.g. color shades of a fish). An example can be seen in the zoom-in presentation of Figure 6.15. Moreover, when there is a false detection and it is placed on the edges of an object (i.e. the zoom-in presentation of Figure 6.16) the proposed Median² may introduce a wrong result. This happens because of the fact that the median value of different candidate patches may not exactly match the true value. In these situations, marine snow removal via inpainting overcomes the difficulty where it provides smoother results and restores the true scene in case of false detections. This is because this approach does not use the neighborhood pixels directly but the learned priors of the whole scene.

Here, the main aim was to eliminate the unwanted signal from video frames as much as possible while the details and structures are restored with high accuracy. Therefore, computational complexity of the algorithm was not considered to be an important factor. As a result, despite the performance improvement of the proposed algorithm using prior knowledge and inpainting, it was noted that the computational complexity of the approach rather increased. This is a consequence of the training phase of statistical priors as well as the iterative scheme of inpainting algorithm. The learning priors is a one-time attempt where the same priors are used for all the test frames of the same video, or in case of an arbitrary training set, the same learned priors can be used for any test image/frame. Thus, one can ignore the time consumption of this stage. Nevertheless, the most time consuming stage of the algorithm is the inpainting step. To give a better insight, for a full HD video on a single 3.4 GHz Core, the background modeling could process around 12 frames per second and the mask extraction takes about 3 – 10 seconds depending on the number of image pyramid layers and the level of corruption with marine snow. However, the Matlab implementation of the inpainting algorithm needs about a minute to restore a frame depending on how corrupted the frame is with marine snow.

6.6 Conclusion

In this chapter, detection, and elimination of marine snow have been addressed. Two scenarios dealing with single images and video frames were considered. This required the detection of particles. Thus, first, characteristics of marine snow in acquired data was studied. This study answers the equations about physical properties of this phenomenon appearing in images or video frames. Next, taking into account these properties, marine snow has been detected. For single images, a novel filtering scheme termed Median² was suggested. The spacial design of this filter, the overlapping scheme and voting algorithm, enabled it to differentiate between object edges and marine snow particles where it also decreased the blurring effect caused due to filtering.

The second approach dealt with marine snow removal where the available data was videos with the static camera where small movements of the camera are allowed. In such an outline, the video frames were used to acquire some statistical priors of the scene that assist inpainting algorithm for removing marine snow by recovering the scene behind the particles. The video frames were employed to model the background image. This image takes into account the temporal coherence of the video and provides information about static components of the scene (the scene covered by

marine snow). The statistical priors of the scene were then learned using this reference image. In employing these priors, together with an inpainting algorithm, the scene behind the marine snow was restored without directly using the neighborhoods. This leads to a more accurate restoration where the uncertainties introduced due to using only neighborhood pixels, were reduced. In addition, the algorithm was adapted to the small movements of the camera such as pan, tilt and zoom. As a result, the algorithm could be used when the camera moves slowly, as long as the scene contains some features to be used for detecting the movement.

Both algorithms were qualitatively evaluated where some simulated frames were also used to give numeric comparisons (PSNR and MSE) of the proposed algorithms to the literature. To provide synthetic data, two approaches were employed. For the first approach, using the same strategy as [BG12] but taking into account different geometries for particles, marine snow was simulated on original UW images. The second strategy suggested a reverse pattern where a test frame containing marine snow was used to provide the ground-truth data. This was done by manually extracting the particles from the frame and refilling the missing pixels using neighborhoods. In this way a ground-truth image corresponding to the current frame was conducted. Next, positioning the sample particles (extracted from the original frame) in the ground-truth image, the simulated data was acquired. With this, the simulated data has the highest correlation to the original one. The presented simulated data in this dissertation were acquired via the manual approach.

The results illustrated the success of the proposed algorithms at detecting and removing marine snow from single images and video frames where the main goal, namely eliminating marine particles when the edges of other objects are preserved, was fulfilled. This conclusion was also proved numerically.

Despite the promising results of the proposed algorithms, there is nevertheless, room for improvement. In this work, the goal was to eliminate marine snow particles, however, when an artificial light at the time of photography is used, the light reflections of marine snow introduce a very complex distortion. These reflections are usually much larger than marine snow itself. In addition, their intensity distribution does not exactly match with one of the particles. The proposed approaches could eliminate some of these reflections, however, it could be very dramatic in some cases where some parts of an image are almost completely covered by this phenomenon, thus, not much information remains in being able to recover the scene behind it.

Moreover, when there are particles of an unusually large size or when several smaller particles form a very large one which can be equal or even larger than some of the objects in the scene, the algorithm cannot detect them.

In both these scenarios, choosing a larger patch size or increasing the number of layers of the image pyramid may increase the chance of detecting the particles, nonetheless, it leads to some false detections of true object edges which is somewhat a bigger disadvantage. Thus, this is a common practice to find a balance between the number of detected particles and the accuracy of the restoration as was done in this piece of work.

7 Combined Super-Resolution, Deblurring and Marine Snow Removal

To obtain optimal results, methods can be combined, this chapter shows how this can be achieved and provides some examples of the result of both algorithms, super-resolution and deblurring, and marine snow removal in one pipeline.

It is not a common event to have an UW image which suffers only from a specific degradation such as blur, haze, color cast or noise, therefore, combining different approaches is beneficial. Marine particles are usually disruptive when they are large enough to cover a scene of interest or distract advance computer vision tasks i.e. segmentation. Thus, having LR data which suffer from strong blur, marine snow could be the least problem. In this case, it is logical to first increase the image resolution and deblur the data. Consequently, if there are marine particles in the scene which are now visible and disruptive, the proposed marine snow detection and removal methods can be advantageous.

However, this is not a rule of thumb so one can decide which approach needs to be applied first depending on the level of distortion. An example could be when the field of view is narrow and the object of interest is very close to the camera. In this case increasing the image resolution and deblurring the image structure is not as critical as removing marine particles, which are now the main problem since they are in focus as the objects of interest in the scene.

Generally, the super-resolution and deblurring algorithm is the first stage. Marine snow detection and removal follows it as the second stage of the pipeline. However, different UW image examples are provided when a different order of pipeline is applied. To this end, both simulated data and real images are employed. For synthetic images, one needs to simulate both UW blur and marine snow. However, due to the fact that UW blur is a function of camera-object distance, a distance map of the scene is needed to simulate this kind of blur. Unfortunately, there is no such data available for UW images. As a result, the LR and blur UW images are employed and marine snow is simulated (as was introduced in the previous chapter). These degraded images are then input to the pipeline.

Figure 7.1 illustrates some examples of synthetic data and the results of the pipeline. For simulated data, since marine snow is more dominant, first marine snow detection and removal is applied. The result of this stage is evaluated qualitatively (Figure 7.1) and quantitatively via PSNR and MSE. Further, the result of this stage is enhanced using the super-resolution and deblurring algorithm. The non-reference metrics, UISM and entropy are applied to assess the output's quality. Table 7.1 demonstrates the numeric results for the simulated data in Figure 7.1.

The results in Table 7.1 are inline with visual comparisons. On examining this table, one can observe that PSNR values of the output images increased noticeably where MSE values confirm the improvement with drastically decreased values. UISM evaluates image sharpness which indicates the success of the deblurring approach and finally, entropy values show improvement in the sense of more available information provided via super-resolution framework.

In addition, the proposed algorithms are evaluated using real images (Figure 7.2). In these images, due to significant blur and low resolution, marine particles are not visible, however, once the image is super-resolved and details are enhanced, marine snow can be observed. Therefore, in such images, first, the input image is enhanced via super-resolution and deblurring, next, the result is processed via marine snow detection and removal algorithm. In this scenario, since there is no reference image available, only non-reference metrics can be used to evaluate the final result. Table 7.1 shows the numeric results of the images in Figure 7.2. As can be observed, the numbers confirm image improvement as regards of image sharpness and entropy.

Table 7.1: Quantitative evaluation of the proposed methods via PSNR, MSE, UISM and Entropy.

		PSNR	MSE	UISM	Entropy
Simulated Data					
img1	Input Image	32.0606	84.7532	1.5980	6.6179
	Output Image	43.2462	6.4506	1.8392	6.8187
img2	Input Image	34.1857	51.9574	1.5609	6.8764
	Output Image	41.9217	8.7508	1.6300	6.9094
img3	Input Image	29.1052	167.3763	1.7331	5.0972
	Output Image	46.3807	3.1344	2.0001	5.2519
Real Data					
img4	Input Image	-	-	2.3588	6.1793
	Output Image	-	-	2.7319	6.3414
img5	Input Image	-	-	5.1910	7.0031
	Output Image	-	-	5.8919	7.1555

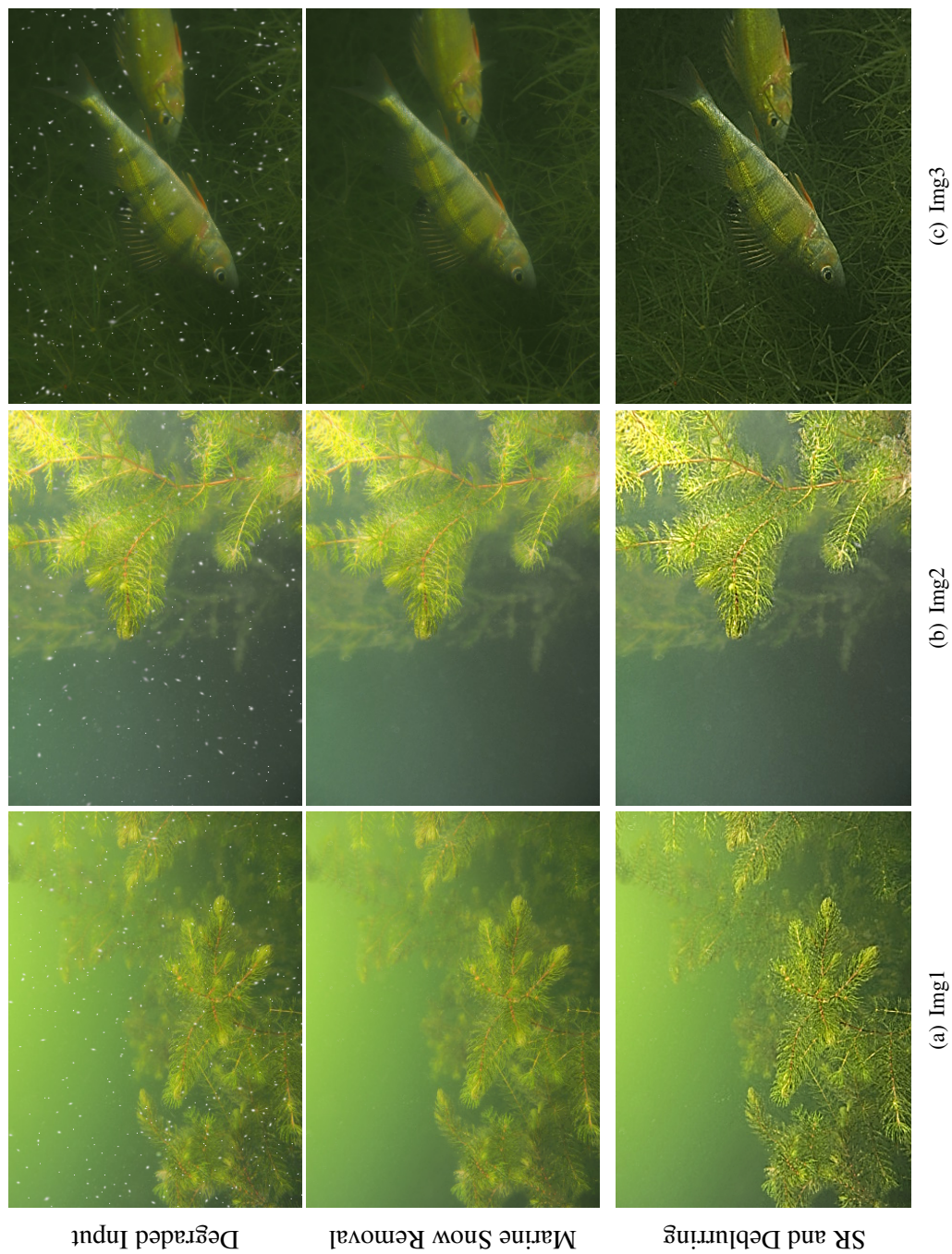


Figure 7.1: Performance of the proposed methods in a pipeline. From top to the bottom: simulated degraded image, marine snow detection and elimination, and finally, super-resolution and deblurring.

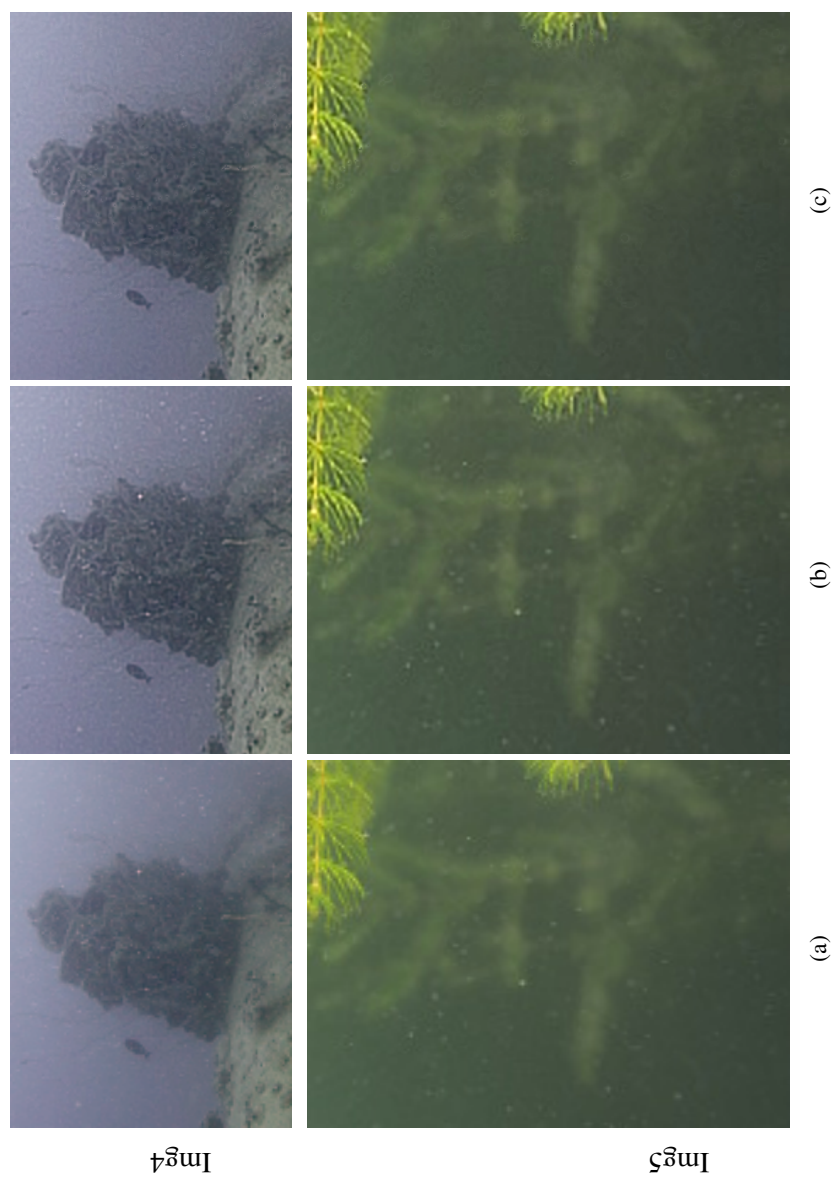


Figure 7.2: Performance of the proposed methods in a pipeline. (a) degraded input image, (b) super-resolution and deblurring, and (c) marine snow detection and elimination.

8 Conclusion and Future Work

8.1 Conclusion

In this thesis, three main objectives were addressed. UW image super-resolution, deblurring and marine snow removal (denoising).

Image super-resolution and deblurring via multiple class dictionaries was applied via sparse representation and dictionary learning over a set of compact class dictionaries. This required a classifier that can cluster the image patches of the same kind, meaning, the patches in each cluster share a common property. Such a classifier needs to be scale invariant since it is used to categorize the training and testing image patches for a super-resolution approach. This implies that the decision made about clustering a specific signal has to be as independent of scale as possible. In considering this, a blurriness measure was developed which is based on the quality of sparse representation over a degraded dictionary. Through several empirical studies, the proposed classifier has shown to be fit for its purposes. A blur degradation model corresponding to the forward scatter component was also integrated into the super-resolution framework to provide UW images with sharp and enhanced details. Applying the proposed sparse coding approach on LR and blurry UW images has demonstrated the-state-of-the-art performance. In summary, the following remarks have been conducted through the super-resolution and deblurring approach.

- A blurriness measure (BM) is proposed as the classifier. The spatial intensity variations (blurriness) of the image patches are the criteria to separate them. The BM is calculated using a dictionary with degraded blurry atoms. Using a fix sparsity, the error in the representation of image patches over this dictionary is the indicator for classification. The logic behind the idea is that a dictionary with blurry atoms has different abilities in representing a blurry or a sharp image patch. Additionally, the proposed classifier has illustrated that it can generally hold for many test image patches with different blurriness levels and is approximately scale-invariant. These findings suggest the usage of the BM for clustering and model selection for the proposed super-resolution and deblurring algorithm.
- Super-resolution via sparse representation over learned compact dictionary classes is shown to achieve promising results where it outperforms the existing state-of-the-art super-resolution algorithms. Furthermore, integrating a deblurring approach into super-resolution has shown a significant improvement for image details and edges. This is especially beneficial for UW images which suffer from spatial variant blur. These achievements are validated through experiments in terms of qualitative and quantitative evaluations.
- In terms of computational complexity, the proposed approach is comparable to one of the leading approaches with a single dictionary (highly redundant) where it even outperforms

the second one. This is on the grounds that using several compact dictionaries is more efficient than using a highly redundant single one.

Denoising: marine snow detection and removal was addressed via two approaches dealing with single images and video frames respectively. For single image marine snow removal, a novel algorithm termed Median² was suggested. The approach detects marine particles according to their physical properties in acquired data. The special design of Median² filter enables it to:

- differentiate between particles and object edges,
- filter the particles larger than single pixels (approximately between 3×3 and 25×25 pixels) without introducing expected blurriness and
- even reduce the blurriness effect introduced by traditional Median filtering.

Despite the promising result, wherever there is too many marine particles which there is not much local information available or whenever the detection of suspended particles is not perfect, the algorithm results in some uncertainty or low accuracy. Thus to overcome such shortcomings and take advantage of extra information provided by video sequence, a second approach dealing with video frames has been posited. The main assumption was that the camera is static, however, small movements of the camera such as pan, tilt, and zoom were allowed. This allows for the slow movements of the robot (i.e. ROV) at the time of acquisition. Dividing the video frames into two groups, training and testing; first, a background model was obtained using the training frames. For this, a mixture of two Gaussians separating the foreground objects from the background was proposed. Moreover, the background modeling algorithm was adapted to some movements of the camera such as pan, tilt and zoom. In order to achieve this, the affine transformation of the movement for the current frame was detected and the background model was adjusted accordingly. Once the background model is obtained, employing this so-called reference image, some statistical priors of the whole scene were learned. This provides information about the structures and their relationships in the scene which are covered by marine snow. Using the proposed Median² filtering, for each test frame, a mask indicating the locations of particles was extracted. Finally, using the extracted mask and an inpainting algorithm, the scene behind marine snow was restored using the learned priors rather than directly using neighborhoods. Consequently, the algorithm:

- achieves a more accurate restoration since the restoration does not use neighborhood but instead uses the statistical features of the very same scene,
- reduces the blurriness even more,
- performs better in the case of any false detections since the image is handled at once rather than patch-based, thus, there is more information available and
- considers the temporal coherence of the video since the background model, learned over training frames of the scene, is used as the training data for learning the priors.

The outcomes of both algorithms were validated via qualitative and quantitative comparisons. It has been shown that the proposed methods could outperform the existing methods where the marine particles are eliminated while the image structures such as object edges and details are preserved as much as possible.

8.2 Future Work

Despite the promising result of the proposed algorithms, there are still a few challenges which can be improved upon:

- The distance map proposed for the super-resolution and deblurring approach is an approximate, which can be affected by significant haze or low contrast.
- The light reflection on the marine particles is still an open issue and cannot be properly detected and removed using the proposed approaches.
- The proposed approaches are designed for accuracy and not for efficiency, thus, one needs to optimize them for real-time performance.
- Having both approaches in one pipeline needs user interaction to define the order of the algorithms.

The following ideas can be made on the conducted work in this thesis as possible future extensions.

Considering the proposed super-resolution and deblurring algorithm:

- The perfect model selection plot (5.10) provided in Chapter 5 indicates the room for improvement. This can be addressed by a more accurate model selection or by combining more than one classifier which can, in turn, reduce the errors in clustering.
- It would be interesting to integrate a haze removal algorithm into the super-resolution and deblurring approach. This way, one can provide a comprehensive approach to UW image degradations. Furthermore, it can improve the model selection procedure when dealing with hazy UW images.

Considering the proposed marine snow detection and removal approaches:

- The proposed methods in Chapter 6 deal with data which are given to the algorithm without having any control over their acquisition. Thus, it would be interesting to investigate the possibilities when one can acquire data in different situations, i.e. different artificial light positionings, its angle, using a polarizer or a stereo system.
- Using artificial light at the time of data acquisition causes an image degradation in the form of extreme light reflection from marine particles. The complexity of this phenomenon and the large size of the defect, makes it very difficult or even impossible to remove. However, one may consider an approach to handle these reflections, such as detecting them via their

8 *Conclusion and Future Work*

geometrical properties, since they mostly appear as circle-shaped defects. Though, in some cases, they are very dominant and there is almost no information available about the scene behind them. In such cases, video sequences or multiple images from the exact same scene may be beneficial.

- Parallelization of training the BG model, prior models and processing steps to achieve real-time performance.

References

- [AE08] M. Aharon and M. Elad. “Sparse and redundant modeling of image content using an image-signature-dictionary”. In: *SIAM Journal on Imaging Sciences* 1.3 (2008), pp. 228–247.
- [AEB06] M. Aharon, M. Elad, and A. Bruckstein. “K -SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation”. In: *IEEE Transactions on Signal Processing* 54.11 (2006), pp. 4311–4322.
- [AMK05a] A. Arnold-Bos, J-P. Malkasse, and G. Kervern. “A preprocessing framework for automatic underwater images denoising”. In: *European Conference on Propagation and Systems*. Brest, France, 2005.
- [AMK05b] A. Arnold-Bos, J. P. Malkasse, and G. Kervern. “Towards a model-free denoising of underwater optical images”. In: *Europe Oceans 2005*. 2005, pp. 527–532.
- [Anc+12] C. Ancuti, C.O. Ancuti, T. Haber, and P. Bekaert. “Enhancing underwater images and videos by fusion”. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2012, pp. 81–88.
- [AS88] A. L. Alldredge and M. W. Silver. “Characteristics, dynamics and significance of marine snow”. In: *Progress in oceanography* 20.1 (1988), pp. 41–82.
- [Ban+14] S. Banerjee, G. Sanyal, S. Ghosh, R. Ray, and S. N. Shome. “Elimination of Marine Snow effect from underwater image - An adaptive probabilistic approach”. In: *IEEE Students’ Conference on Electrical, Electronics and Computer Science (SCEECS)* (2014), pp. 1–4.
- [Baz+06] S. Bazeille, I. Quidu, L. Jaulin, and J-P. Malkasse. “Automatic underwater image pre-processing”. In: *CMM’06*. 2006.
- [BCM05] A. Buades, B. Coll, and J-M. Morel. “A non-local algorithm for image denoising”. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*. Vol. 2. 2005, pp. 60–65.
- [BDE09] A. M. Bruckstein, D. L. Donoho, and M. Elad. “From sparse solutions of systems of equations to sparse modeling of signals and images”. In: *SIAM review* 51.1 (2009), pp. 34–81.
- [BE08] O. Bryt and M. Elad. “Compression of facial images using the K-SVD algorithm”. In: *Journal of Visual Communication and Image Representation* 19.4 (2008), pp. 270–282.
- [Ber+08] R. Bergman, R. Maurer, H. Nachlieli, G. Ruckenstein, P. Chase, and D. Greig. “Comprehensive solutions for automatic removal of dust and scratches from images”. In: *Journal of Electronic Imaging* 17.1 (2008).

References

- [BG12] M. Boffety and F. Galland. “Phenomenological marine snow model for optical underwater image simulation: Applications to color restoration”. In: *OCEANS, IEEE*. 2012, pp. 1–6.
- [Bor+99] M. Born et al. *Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light*. 7th ed. Cambridge University Press, 1999.
- [BR11] B. Benfold and I. Reid. “Stable Multi-Target Tracking in Real-Time Surveillance Video”. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*. 2011, pp. 3457–3464.
- [CB00] C. Couvreur and Y. Bresler. “On the optimality of the backward greedy algorithm for the subset selection problem”. In: *SIAM Journal on Matrix Analysis and Applications* 21.3 (2000), pp. 797–808.
- [CBL89] S. Chen, S. A. Billings, and W. Luo. “Orthogonal least squares methods and their application to non-linear system identification”. In: *International Journal of control* 50.5 (1989), pp. 1873–1896.
- [CC12] J.Y. Chiang and Ying-Ching Chen. “Underwater Image Enhancement by Wavelength Compensation and Dehazing”. In: *IEEE Transactions on Image Processing* 21.4 (2012), pp. 1756–1769.
- [CDS01] S. S. Chen, D. L. Donoho, and M. A. Saunders. “Atomic decomposition by basis pursuit”. In: *SIAM review* 43.1 (2001), pp. 129–159.
- [Cha+03] M. Chambah, D. Semani, A. Renouf, P. Courtellemont, and A. Rizzi. “Underwater color constancy: enhancement of automatic live fish recognition”. In: *Electronic Imaging 2004*. 2003, pp. 157–168.
- [Che+11] Y. Chen, W. Li, M. Xia, Q. Li, and K. Yang. “Super-resolution reconstruction for underwater imaging”. In: *Optica Applicata* 41.4 (2011), pp. 841–853.
- [CHN04] R. H. Chan, C. Hu, and M. Nikolova. “An iterative procedure for removing random-valued impulse noise”. In: *IEEE signal processing letters* (2004), pp. 921–924.
- [CME10] N. Carlevaris-Bianco, A. Mohan, and R.M. Eustice. “Initial results in underwater single image dehazing”. In: *OCEANS, IEEE*. 2010, pp. 1–8.
- [CR02] S.F. Cotter and B.D. Rao. “Application of total least squares (TLS) to the design of sparse signal representation dictionaries”. In: *Conference Record of the Thirty-Sixth Asilomar Conference on Signals, Systems and Computers*. Vol. 1. 2002, pp. 963–966.
- [CT05] P. Choudhury and J. Tumblin. “The trilateral filter for high contrast images and meshes”. In: *ACM SIGGRAPH 2005 Courses*. 2005, p. 5.
- [CW01] T. Chen and H. R. Wu. “Adaptive impulse detection using center-weighted median filters”. In: *IEEE Signal Processing Letters* (2001), pp. 1–3.
- [CYV00] S. G. Chang, B. Yu, and M. Vetterli. “Adaptive wavelet thresholding for image denoising and compression”. In: *IEEE transactions on image processing* 9.9 (2000), pp. 1532–1546.

- [CYX04] H. Chang, D-Y. Yeung, and Y. Xiong. “Super-resolution through neighbor embedding”. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*. Vol. 1. 2004.
- [Dab+08] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. “Image restoration by sparse 3D transform-domain collaborative filtering”. In: *Electronic Imaging*. 2008.
- [DBP57] S. Q. Duntley, A. R. Boileau, and R. W. Preisendorfer. “Image transmission by the troposphere I”. In: *JOSA* 47.6 (1957), pp. 499–506.
- [DH01] D. L. Donoho and X. Huo. “Uncertainty principles and ideal atomic decomposition”. In: *IEEE Transactions on Information Theory* 47.7 (2001), pp. 2845–2862.
- [Dix+83] T. H. Dixon, T.J. Pivrotto, R.F. Chapman, and R.C. Tyce. “A range-gated laser system for ocean floor imaging”. In: *Marine Technology Society Journal* 17.4 (1983), pp. 5–12.
- [DMA97] G.f Davis, S. Mallat, and M. Avellaneda. “Adaptive greedy approximations”. In: *Constructive approximation* 13.1 (1997), pp. 57–98.
- [Don+11] W. Dong, L. Zhang, G. Shi, and X. Wu. “Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization”. In: *IEEE Transactions on Image Processing* 20.7 (2011), pp. 1838–1857.
- [Don06] D. L. Donoho. “For most large underdetermined systems of linear equations the minimal ℓ_1 -norm solution is also the sparsest solution”. In: *Communications on pure and applied mathematics* 59.6 (2006), pp. 797–829.
- [Dun63] S. Q. Duntley. “Light in the Sea*”. In: *JOSA* 53.2 (1963), pp. 214–233.
- [EA06] M. Elad and M. Aharon. “Image denoising via learned dictionaries and sparse representation”. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Vol. 1. 2006, pp. 895–900.
- [EAH99] K. Engan, S. O. Aase, and J. H. Husoy. “Method of optimal directions for frame design”. In: *IEEE International Conference on Acoustics, Speech, and Signal Processing, 1999. Proceedings*. Vol. 5. 1999, pp. 2443–2446.
- [EB02] M. Elad and A. M. Bruckstein. “A generalized uncertainty principle and sparse representation in pairs of bases”. In: *IEEE Transactions on Information Theory* 48.9 (2002), pp. 2558–2567.
- [Ela+05] M. Elad, J-L Starck, P. Querre, and D. L. Donoho. “Simultaneous cartoon and texture image inpainting using morphological component analysis (MCA)”. In: *Applied and Computational Harmonic Analysis* 19.3 (2005), pp. 340–358.
- [Eng99] K. Engan. “Frame design using FOCUSS with method of optimal directions (MOD)”. In: *Journal of NORSIG*. 1999.
- [ERK00] K. Engan, B. D. Rao, and K. Kreutz-Delgado. “Regularized FOCUSS for subset selection in noise”. In: *Journal of NORSIG*. 2000, pp. 247–250.
- [ESH07] K. Engan, K. Skretting, and J. H. Husøy. “Family of iterative LS-based dictionary learning algorithms, ILS-DLA, for sparse signal representation”. In: *Digital Signal Processing* 17.1 (2007), pp. 32–49.

References

- [Fan+13] S. Fang, R. Deng, Y. Cao, and C. Fang. “Effective single underwater image enhancement by fusion”. In: *Journal of computers* 8.4 (2013), pp. 904–911.
- [Fen+11] J. Feng, L. Song, X. Yang, and W. Zhang. “Learning dictionary via subspace segmentation for sparse representation”. In: *18th IEEE International Conference on Image Processing (ICIP)*. 2011, pp. 1245–1248.
- [Fer+06] R. Fergus, B. Singh, A. Hertzmann, S. T. Roweis, and W. T. Freeman. “Removing camera shake from a single photograph”. In: *ACM Transactions on Graphics (TOG)*. Vol. 25. 3. 2006, pp. 787–794.
- [FS03] P. Favaro and S. Soatto. “Seeing beyond occlusions (and other marvels of a finite lens aperture)”. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Vol. 2. 2003, pp. 579–586.
- [FS05] P. Favaro and S. Soatto. “A geometric approach to shape from defocus”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27.3 (2005), pp. 406–417.
- [FXG11] S. Feifei, Z. Xuemeng, and W. Guoyu. “An approach for underwater image denoising via wavelet decomposition and high-pass filter”. In: *IEEE International Conference on Intelligent Computation Technology and Automation (ICICTA)*. Vol. 2. 2011, pp. 417–420.
- [GG84] S. Geman and D. Geman. “Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images”. In: *IEEE Transactions on pattern analysis and machine intelligence* 6 (1984), pp. 721–741.
- [GI14] A. S. A. Ghani and N. A. M. Isa. “Underwater image quality enhancement through composition of dual-intensity images and Rayleigh-stretching”. In: *SpringerPlus* 3.1 (2014), p. 757.
- [GLW16] Y. Gao, H. Li, and S. Wen. “Restoration and Enhancement of Underwater Images Based on Bright Channel Prior”. In: *Journal on Mathematical Problems in Engineering* (2016).
- [GMS03] A. C. Gilbert, S. Muthukrishnan, and M. J. Strauss. “Approximation of functions over redundant dictionaries using coherence”. In: *Proceedings of the fourteenth annual ACM-SIAM symposium on Discrete algorithms*. 2003, pp. 243–252.
- [GR92] D. Geman and G. Reynolds. “Constrained restoration and the recovery of discontinuities”. In: *IEEE Transactions on pattern analysis and machine intelligence* 14.3 (1992), pp. 367–383.
- [GR97] I. F. Gorodnitsky and B. D. Rao. “Sparse signal reconstruction from limited data using FOCUSS: A re-weighted minimum norm algorithm”. In: *IEEE Transactions on signal processing* 45.3 (1997), pp. 600–616.
- [Gul06] O. G. Guleryuz. “Nonlinear approximation based image recovery using adaptive sparse reconstructions and iterated denoising-part I: theory”. In: *IEEE Transactions on image processing* 15.3 (2006), pp. 539–554.

- [Gut+10] E. Gutzeit, S. Ohl, A. Kuijper, J. Voskamp, and B. Urban. “Setting Graph Cut Weights for Automatic Foreground Extraction in Wood Log Images”. In: *International Conference on Computer Vision Theory and Applications (VISAPP)*. 2010, pp. 60–67.
- [Hah06] D. W. Hahn. “Light scattering theory”. In: *Department of Mechanical and Aerospace Engineering, Florida* (2006).
- [HH95] H. Hwang and R. A. Haddad. “Adaptive median filters: new algorithms and results”. In: *IEEE Transactions on Image Processing* (1995), pp. 499–502.
- [Hin99] Geoffrey E Hinton. “Products of experts”. In: *Proceedings of the Ninth International Conference on Artificial Neural Networks (ICANN)* 1 (1999), pp. 1–6.
- [Hit+13] M. S. Hitam, W. N. J. H. W. Yussof, E. Afreen Awalludin, and Z. Bachok. “Mixture contrast limited adaptive histogram equalization for underwater image enhancement”. In: *IEEE International Conference on Computer Applications Technology (ICCAT)*. 2013, pp. 1–5.
- [Hou+07a] W. Hou, D.J. Gray, A.D. Weidemann, G.R. Fournier, and J.L. Forand. “Automated underwater image restoration and retrieval of related optical properties”. In: *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. 2007, pp. 1889–1892.
- [Hou+07b] W. Hou, A. D. Weidemann, D. J. Gray, and G. R. Fournier. “Imagery-derived modulation transfer function and its applications for underwater imaging”. In: *Application of Digital Image Processing (SPIE)*. Vol. 6696. 2007, pp. 22–28.
- [Hou+08] W. Hou, D. J. Gray, A. D. Weidemann, and R. A. Arnone. “Comparison and validation of point spread models for imaging in natural waters”. In: *Optics Express* 16.13 (2008), pp. 9958–9965.
- [HS08] Y. Hel-Or and D. Shaked. “A discriminative approach for wavelet denoising”. In: *IEEE transactions on image processing* 17.4 (2008), pp. 443–457.
- [HSA15] J-B. Huang, A. Singh, and N. Ahuja. “Single image super-resolution from transformed self-exemplars”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2015, pp. 5197–5206.
- [HST09] K. He, J. Sun, and X. Tang. “Single image haze removal using dark channel prior”. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2009, pp. 1956–1963.
- [HST11] K. He, J. Sun, and X. Tang. “Single image haze removal using dark channel prior”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33.12 (2011), pp. 2341–2353.
- [HST13] K. He, J. Sun, and X. Tang. “Guided image filtering”. In: *IEEE transactions on pattern analysis and machine intelligence* 35.6 (2013), pp. 1397–1409.
- [HW07] W. Hou and A. D. Weidemann. “Objectively assessing underwater image quality for the purpose of automated restoration”. In: *Visual Information Processing XVI*. Vol. 6575. 2007.

References

- [ICG86] D. S. Immel, M. F. Cohen, and D. P. Greenberg. "A Radiosity Method for Non-diffuse Environments". In: *Proceedings of the 13th Annual Conference on Computer Graphics and Interactive Techniques*. 1986, pp. 133–142.
- [IP91] M. Irani and S. Peleg. "Improving resolution by image registration". In: *Graphical models and image processing (CVGIP)* 53.3 (1991), pp. 231–239.
- [Iqb+07] K. Iqbal, R. Abdul S., M. Osman, A. Z. Talib, et al. "Underwater Image Enhancement Using An Integrated Colour Model." In: *IAENG International Journal of Computer Science* 32.2 (2007), pp. 239–244.
- [Iqb+10] K. Iqbal, M. Odetayo, A. James, R.A. Salam, and A.Z.H. Talib. "Enhancing the low quality images using Unsupervised Colour Correction Method". In: *IEEE International Conference on Systems Man and Cybernetics (SMC)*. 2010, pp. 1703–1709.
- [Jaf90] J.S. Jaffe. "Computer modeling and the design of optimal underwater imaging systems". In: *IEEE Journal of Oceanic Engineering* 15.2 (1990), pp. 101–111.
- [JG12] D. x. Jia and Y. r. Ge. "Underwater image de-noising algorithm based on Nonsub-sampled Contourlet Transform and Total Variation". In: *International Conference on Computer Science and Information Processing (CSIP)*. 2012, pp. 76–80.
- [JRW97] D. J. Jobson, Z. Rahman, and G. A. Woodell. "A multiscale retinex for bridging the gap between color images and the human observation of scenes". In: *IEEE Transactions on Image Processing* 6.7 (1997), pp. 965–976.
- [Kaj86] J. T. Kajiya. "The Rendering Equation". In: *SIGGRAPH Comput. Graph.* 20.4 (1986), pp. 143–150.
- [KGC02] L. Kaur, S. Gupta, and RC. Chauhan. "Image Denoising Using Wavelet Thresholding." In: *International Conference ICVGIP*. Vol. 2. 2002, pp. 16–18.
- [Kha+16] A. Khan, S. S. A. Ali, A. S. Malik, A. Anwer, and F. Meriaudeau. "Underwater image enhancement by wavelet based fusion". In: *IEEE International Conference on Underwater System Technology: Theory and Applications (USYS)*. 2016, pp. 83–88.
- [Kim+13] J-H. Kim, W-D. Jang, J-Y. Sim, and C-S. Kim. "Optimized contrast enhancement for real-time image and video dehazing". In: *Journal of Visual Communication and Image Representation* 24.3 (2013), pp. 410–425.
- [Kos24] H. Koschmieder. *Theorie der horizontalen Sichtweite: Kontrast und Sichtweite*. Beiträge zur Physik der freien Atmosphäre v. 2. 1924.
- [Kov99] P. Kovsi. "Phase preserving denoising of images". In: *Australian Patt. Recog. Soc. Conf. DICTA* 4.3 (1999), pp. 212–217.
- [Kre+03] K. Kreutz-Delgado, J. F. Murray, B. D. Rao, K. Engan, T-W. Lee, and T. J. Sejnowski. "Dictionary learning algorithms for sparse representation". In: *Neural computation* 15.2 (2003), pp. 349–396.

- [LC16] H. Liu and L. P. Chau. “Underwater image restoration based on contrast enhancement”. In: *2016 IEEE International Conference on Digital Signal Processing (DSP)*. 2016, pp. 584–588.
- [Les+05] S. Lesage, R. Gribonval, Fr. Bimbot, and L. Benaroya. “Learning unions of orthonormal bases with thresholded singular value decomposition”. In: *IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005. Proceedings.(ICASSP’05)*. Vol. 5. 2005.
- [LH02] B Lee and M Hedley. “Background estimation for video surveillance”. In: *Image and vision computing New Zealand*. 2002, pp. 315–320.
- [Li+16] C-Y. Li, J-C. Guo, R-M. Cong, Y-W. Pang, and B. Wang. “Underwater Image Enhancement by Dehazing With Minimum Information Loss and Histogram Distribution Prior”. In: *IEEE Transactions on Image Processing* 25.12 (2016), pp. 5664–5677.
- [Liu+01] Z. Liu, Y. Yu, K. Zhang, and H. Huang. “Underwater image transmission and blurred image restoration”. In: *Optical Engineering* 40.6 (2001), pp. 1125–1131.
- [Liu+15a] X. Liu, R. Nian, B. He, and A. Lendasse. “A rapid weighted median filter based on saliency region for underwater image denoising”. In: *OCEANS, IEEE*. 2015, pp. 1–4.
- [Liu+15b] X Liu, R. Nian, B. He, and A. Lendasse. “A rapid weighted median filter based on saliency region for underwater image denoising”. In: *OCEANS, IEEE*. 2015, pp. 1–4.
- [LLS13] H. Lu, Y. Li, and S. Serikawa. “Underwater image enhancement using guided trigonometric bilateral filter and fast automatic color correction”. In: *20th IEEE International Conference on Image Processing (ICIP)*. 2013, pp. 3412–3416.
- [LLW08] A. Levin, D. Lischinski, and Y. Weiss. “A closed-form solution to natural image matting”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 30.2 (2008), pp. 228–242.
- [LM71] E. H. Land and J. J. McCann. “Lightness and retinex theory”. In: *Josa* 61.1 (1971), pp. 1–11.
- [LO99] M. S Lewicki and B. A. Olshausen. “Probabilistic framework for the adaptation and comparison of image codes”. In: *JOSA A* 16.7 (1999), pp. 1587–1601.
- [Lu+15] H. Lu, Y. Li, L. Zhang, and S. Serikawa. “Contrast enhancement for images in turbid water”. In: *JOSA A* 32.5 (2015), pp. 886–893.
- [Lu+17] Huimin Lu, Yujie Li, Shota Nakashima, Hyongseop Kim, and Seiichi Serikawa. “Underwater image super-resolution by descattering and fusion”. In: *IEEE Access* 5 (2017), pp. 670–679.
- [Mai+08] B. Mailhé, S. Lesage, R. Gribonval, F. Bimbot, and P. Vanderghelynst. “Shift-invariant dictionary learning for sparse representations: extending K-SVD”. In: *16th IEEE European Signal Processing Conference*. 2008, pp. 1–5.

References

- [Mai+09a] J. Mairal, F. Bach, J. Ponce, and G. Sapiro. “Online dictionary learning for sparse coding”. In: *Proceedings of the 26th annual international conference on machine learning*. 2009, pp. 689–696.
- [Mai+09b] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman. “Non-local sparse models for image restoration”. In: *IEEE 12th International Conference on Computer Vision*. 2009, pp. 2272–2279.
- [Mai+10] J. Mairal, F. Bach, J. Ponce, and G. Sapiro. “Online learning for matrix factorization and sparse coding”. In: *Journal of Machine Learning Research* 11 (2010), pp. 19–60.
- [McG80] B. L. McGlamery. “A Computer Model For Underwater Camera Systems”. In: *Ocean Optics VI*. Vol. 208. 1980, pp. 221–232.
- [MES08] J. Mairal, M. Elad, and G. Sapiro. “Sparse representation for color image restoration”. In: *IEEE Transactions on image processing* 17.1 (2008), pp. 53–69.
- [MMB12] A. Mittal, A. K. Moorthy, and A. C. Bovik. “No-reference image quality assessment in the spatial domain”. In: *IEEE Transactions on Image Processing* 21.12 (2012), pp. 4695–4708.
- [Mob94] C. D. Mobley. “Light and water: Radiative transfer in natural waters”. In: *Academic Press* (1994).
- [MS95] N. JB. McFarlane and C. P. Schofield. “Segmentation and tracking of piglets in images”. In: *Machine vision and applications* 8.3 (1995), pp. 187–193.
- [MZ93] S. G. Mallat and Z. Zhang. “Matching pursuits with time-frequency dictionaries”. In: *IEEE Transactions on signal processing* 41.12 (1993), pp. 3397–3415.
- [OF96] B. A. Olshausen and D. J. Field. “Natural image statistics and efficient coding”. In: *Network: computation in neural systems* 7.2 (1996), pp. 333–339.
- [OF97] B. A. Olshausen and D. J. Field. “Sparse coding with an overcomplete basis set: A strategy employed by V1?” In: *Vision research* 37.23 (1997), pp. 3311–3325.
- [Pan+11] K. Panetta, S. Agaian, Y. Zhou, and E. J. Wharton. “Parameterized logarithmic framework for image enhancement”. In: *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 41.2 (2011), pp. 460–473.
- [PC17] Y-T. Peng and P. C. Cosman. “Underwater Image Restoration Based on Image Blurriness and Light Absorption”. In: *IEEE Transactions on Image Processing* 26.4 (2017), pp. 1579–1594.
- [PCC09] F. Petit, A. Capelle-Laize, and P. Carré. “Underwater image enhancement by attenuation inversion with quaternions”. In: *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2009, pp. 1177–1180.
- [PE09] M. Protter and M. Elad. “Image sequence denoising via sparse and redundant representations”. In: *IEEE Transactions on Image Processing* 18.1 (2009), pp. 27–35.

- [PFS10] G. Peyré, J. Fadili, and J-L. Starck. “Learning the morphological diversity”. In: *SIAM Journal on Imaging Sciences* 3.3 (2010), pp. 646–669.
- [PGA16a] K. Panetta, C. Gao, and S. Agaian. “Human-Visual-System-Inspired Underwater Image Quality Measures”. In: *IEEE Journal of Oceanic Engineering* 41.3 (2016), pp. 541–551.
- [PGA16b] K. Panetta, C. Gao, and S. Agaian. “Human-visual-system-inspired underwater image quality measures”. In: *IEEE Journal of Oceanic Engineering* 41.3 (2016), pp. 541–551.
- [Pis+98] E. D. Pisano et al. “Contrast limited adaptive histogram equalization image processing to improve the detection of simulated spiculations in dense mammograms”. In: *Journal of Digital imaging* 11.4 (1998), pp. 193–200.
- [PK10] C.J. Prabhakar and PU P. Kumar. “Underwater image denoising using adaptive wavelet subband thresholding”. In: *IEEE International Conference on Signal and Image Processing (ICSIP)*. 2010, pp. 322–327.
- [PK12] C. J. Prabhakar and P. U. Praveen Kumar. “An Image Based Technique for Enhancement of Underwater Images”. In: *arXiv preprint arXiv:1212.0291* (2012).
- [PM90] P. Perona and J. Malik. “Scale-space and edge detection using anisotropic diffusion”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 12.7 (1990), pp. 629–639.
- [PRK93] Y. C. Pati, R. Rezaifar, and PS. Krishnaprasad. “Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition”. In: *IEEE Conference Record of The Twenty-Seventh Asilomar Conference on Signals, Systems and Computers*. 1993, pp. 40–44.
- [PSA14] K. Panetta, A. Samani, and S. Agaian. “Choosing the optimal spatial domain measure of enhancement for mammogram images”. In: *Journal of Biomedical Imaging* 2014 (2014), p. 3.
- [PZC15] Y-T. Peng, X. Zhao, and P. C. Cosman. “Single underwater image enhancement using depth estimation based on blurriness”. In: *2015 IEEE International Conference on Image Processing (ICIP)*. 2015, pp. 4952–4956.
- [RB05] S. Roth and M. J. Black. “Fields of experts: A framework for learning image priors”. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 2005, pp. 860–867.
- [RG04] G. Rath and C. Guillemot. “Subspace-based error and erasure correction with DFT codes for wireless channels”. In: *IEEE Transactions on Signal Processing* 52.11 (2004), pp. 3241–3252.
- [RG15] M. Radolko and E. Gutzeit. “Video Segmentation via a Gaussian Switch Background Model and Higher Order Markov Random Fields”. In: *Proceedings of the 10th International Conference on Computer Vision Theory and Applications*. Vol. 1. 2015, pp. 537–544.

References

- [RGM03] A. Rizzi, C. Gatta, and D. Marini. “A New Algorithm for Unsupervised Global and Local Color Correction”. In: *Pattern Recogn. Lett.* 24.11 (2003), pp. 1663–1677.
- [RL02] L. Rebollo-Neira and D. Lowe. “Optimized orthogonal matching pursuit approach”. In: *IEEE Signal Processing Letters* 9.4 (2002), pp. 137–140.
- [ROF92] L. I. Rudin, S. Osher, and E. Fatemi. “Nonlinear total variation based noise removal algorithms”. In: *Physica D: Nonlinear Phenomena* 60.1-4 (1992), pp. 259–268.
- [San02] S. J. Sangwine. “Mathematical approaches to linear vector filtering of color images”. In: *Conference on Colour in Graphics, Imaging, and Vision*. Vol. 2002. 1. 2002, pp. 348–351.
- [SB14] T. Stephan and J. Beyerer. “Computergraphical Model for Underwater Image Simulation and Restoration”. In: *Workshop on Computer Vision for Analysis of Underwater Imagery (CVAUI)*. 2014, pp. 73–79.
- [SBR11] W. H. Slade, E. Boss, and C. Russo. “Effects of particle aggregation and disaggregation on their inherent optical properties”. In: *Optics express* 19.9 (2011), pp. 7945–7959.
- [SCE01] A. Skodras, C. Christopoulos, and T. Ebrahimi. “The JPEG 2000 still image compression standard”. In: *IEEE Signal processing magazine* 18.5 (2001), pp. 36–58.
- [SE07] K. Srinivasan and D. Ebenezer. “A new fast and efficient decision-based algorithm for removal of high-density impulse noises”. In: *IEEE signal processing letters* (2007), pp. 189–192.
- [SE10] K. Skretting and K. Engan. “Recursive least squares dictionary learning algorithm”. In: *IEEE Transactions on Signal Processing* 58.4 (2010), pp. 2121–2130.
- [SED04] J-L. Starck, M. Elad, and D. L. Donoho. “Redundant multiscale transforms and their application for morphological component separation”. In: *Advances in Imaging and Electron Physics* 132 (2004), pp. 287–348.
- [SED05] J-L Starck, M. Elad, and D. L. Donoho. “Image decomposition via the combination of sparse representations and a variational approach”. In: *IEEE transactions on image processing* 14.10 (2005), pp. 1570–1582.
- [SG99] C. Stauffer and W.E.L. Grimson. “Adaptive background mixture models for real-time tracking”. In: *Proceedings 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Vol. 2. 1999, pp. 246–252.
- [SHG08] O. G. Sezer, O. Harmanci, and O. G. Guleryuz. “Sparse orthonormal transforms for image compression”. In: *15th IEEE International Conference on Image Processing (ICIP)*. 2008, pp. 149–152.
- [SJA08] Q. Shan, J. Jia, and A. Agarwala. “High-quality motion deblurring from a single image”. In: *ACM Transactions on Graphics (SIGGRAPH)*. Vol. 27. 3. 2008.
- [SK05] Y.Y. Schechner and N. Karpel. “Recovery of underwater visibility and structure by polarization analysis”. In: *IEEE Journal of Oceanic Engineering* 30.3 (2005), pp. 570–587.

- [SK53] N. Suzuki and K. Kato. “Studies on suspended materials marine snow in the sea: part I. sources of marine snow”. In: *Bulletin of the faculty of fisheries hokkaido university* 4.2 (1953), pp. 132–137.
- [SL14] S. Serikawa and H. Lu. “Underwater Image Dehazing Using Joint Trilateral Filter”. In: *Comput. Electr. Eng.* 40.1 (2014), pp. 41–50.
- [SNN01] Y.Y. Schechner, S.G. Narasimhan, and S.K. Nayar. “Instant dehazing of images using polarization”. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*. Vol. 1. 2001, pp. 325–332.
- [SSS07] S. Sudha, G.R. Suresh, and R. Sukanesh. “Wavelet based image denoising using adaptive subband thresholding”. In: *International Journal of Soft Computing* 2.5 (2007), pp. 628–632.
- [SSS13] M. Shanmugasundaram, S. Sukumaran, and N. Shanmugavadivu. “Fusion based denoise-engine for underwater images using curvelet transform”. In: *International Conference on Advances in Computing, Communications and Informatics (ICACCI)* (2013), pp. 941–946.
- [Sun+03] J. Sun, N-N. Zheng, H. Tao, and H-Y. Shum. “Image hallucination with primal sketch priors”. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*. Vol. 2. 2003.
- [SXJ15] J. Shi, L. Xu, and J. Jia. “Just noticeable defocus blur detection and estimation”. In: *IEEE Conference on Computer Vision and Pattern Recognition*. 2015, pp. 657–665.
- [TD05] L. A. Torres-Méndez and G. Dudek. “Color Correction of Underwater Images for Aquatic Robot Inspection”. In: *Proceedings of the 5th International Conference on Energy Minimization Methods in Computer Vision and Pattern Recognition*. 2005, pp. 60–73.
- [TM98] C. Tomasi and R. Manduchi. “Bilateral filtering for gray and color images”. In: *Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271)*. 1998, pp. 839–846.
- [TO06] E. Trucco and A.T. Olmos-Antillon. “Self-Tuning Underwater Image Restoration”. In: *IEEE Journal of Oceanic Engineering* 31.2 (2006), pp. 511–519.
- [TS09] T. Treibitz and Y.Y. Schechner. “Active Polarization Descattering”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 31.3 (2009), pp. 385–399.
- [TYW14] K. Tang, J. Yang, and J. Wang. “Investigating haze-relevant features in a learning framework for image dehazing”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2014, pp. 2995–3000.
- [Wan+12] S. Wang, L. Zhang, Y. Liang, and Q. Pan. “Semi-coupled dictionary learning with applications to image super-resolution and photo-sketch synthesis”. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2012, pp. 2216–2223.

References

- [Wan+14] R. Wang, F. Bunyak, G. Seetharaman, and K. Palaniappan. “Static and Moving Object Detection Using Flux Tensor with Split Gaussian Models”. In: *2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 2014, pp. 420–424.
- [Wel69] W. H. Wells. “Loss of resolution in water as a result of multiple small-angle scattering”. In: *JOSA* 59.6 (1969), pp. 686–691.
- [Wel73] W. H. Wells. “Theory of small angle scattering”. In: *Optics of the Sea* 61 (1973).
- [WL13] X. Wu and H. Li. “A simple and comprehensive model for underwater image restoration”. In: *IEEE International Conference on Information and Automation (ICIA)*. 2013, pp. 699–704.
- [Wre+97] C. Wren, A. Azarbayejani, T. Darrell, and A. Pentland. “Pfindex: Real-Time Tracking of the Human Body”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19 (1997), pp. 780–785.
- [WZ99] Z. Wang and D. Zhang. “Progressive switching median filter for the removal of impulse noise from highly corrupted images”. In: *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing* (1999), pp. 78–80.
- [WZF11] G. Wang, B. Zheng, and F. Fei Sun. “Estimation-based approach for underwater image restoration”. In: *Optics letters* 36.13 (2011), pp. 2384–2386.
- [Yan+08] J. Yang, J. Wright, T. Huang, and Yi-Ma. “Image super-resolution as sparse representation of raw image patches”. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2008, pp. 1–8.
- [Yan+10] J. Yang, J. Wright, T. S. Huang, and Y. Ma. “Image super-resolution via sparse representation”. In: *IEEE transactions on image processing* 19.11 (2010), pp. 2861–2873.
- [Yan+11] H-Y. Yang, P-Y. Chen, C-C. Huang, Y-Z. Zhuang, and Y-H. Shiau. “Low Complexity Underwater Image Enhancement Based on Dark Channel Prior”. In: *Second International Conference on Innovations in Bio-inspired Computing and Applications (IBICA)*. 2011, pp. 17–20.
- [Yan+12a] J. Yang, Z. Wang, Z. Lin, X. Shu, and T. Huang. “Bilevel sparse coding for coupled feature spaces”. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2012, pp. 2360–2367.
- [Yan+12b] S. Yang, M. Wang, Y. Chen, and Y. Sun. “Single-image super-resolution reconstruction via learned geometric dictionaries and clustered sparse coding”. In: *IEEE Transactions on Image Processing* 21.9 (2012), pp. 4016–4028.
- [YBD09] M. Yaghoobi, T. Blumensath, and M. E. Davies. “Dictionary learning for sparse approximations with the majorization method”. In: *IEEE Transactions on Signal Processing* 57.6 (2009), pp. 2178–2191.
- [YS15] M. Yang and A. Sowmya. “An underwater color image quality evaluation metric”. In: *IEEE Transactions on Image Processing* 24.12 (2015), pp. 6062–6071.

- [YSM10] G. Yu, G. Sapiro, and S. Mallat. “Image modeling and enhancement via structured sparse model selection”. In: *17th IEEE International Conference on Image Processing (ICIP)*. 2010, pp. 1641–1644.
- [Yus+13] W. N. J. H. W. Yussof, M. S. Hitam, E. Afreen Awalludin, and Z. Bachok. “Performing Contrast Limited Adaptive Histogram Equalization Technique on Combined Color Models for Underwater Image Enhancement”. In: *International Journal of Interactive Digital Media* 1.1 (2013), pp. 1–6.
- [ZEP10] R. Zeyde, M. Elad, and M. Protter. “On single image scale-up using sparse-representations”. In: *International conference on curves and surfaces*. 2010, pp. 711–730.
- [Zha+17] S. Zhang, T. Wang, J. Dong, and H. Yu. “Underwater image enhancement via extended multi-scale retinex”. In: *Neurocomputing* (2017).
- [Zhe+06] J. Zheng, Y. Wang, N. Nihan, and M. Hallenbeck. “Extracting roadway background image: Mode-based approach”. In: *Transportation Research Record: Journal of the Transportation Research Board* 1944 (2006), pp. 82–88.
- [Zim+88] J. B. Zimmerman, S. M. Pizer, E. V. Staab, J. R. Perry, W. McCartney, and B. C. Brenton. “An evaluation of the effectiveness of adaptive histogram equalization for contrast enhancement”. In: *IEEE Transactions on Medical Imaging* 7.4 (1988), pp. 304–312.
- [ZJQ15] X. Zhao, T. Jin, and S. Qu. “Deriving inherent optical properties from background color and underwater image enhancement”. In: *Ocean Engineering* 94 (2015), pp. 163–172.
- [ZL06] J. Zepeda and F. Labeau. “Tandem Filter Bank-DFT Code for Bursty Erasure Correction”. In: *64th IEEE Vehicular Technology Conference*. 2006, pp. 1–5.
- [ZLL07] Y. Zhao, D. Li, and Z. Li. “Performance enhancement and analysis of an adaptive median filter”. In: *IEEE Second International Conference on Communications and Networking in China*. 2007, pp. 651–653.
- [ZM97] S. C. Zhu and D. Mumford. “Prior learning and Gibbs reaction-diffusion”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19.11 (1997), pp. 1236–1250.
- [ZTG03] L. Zhishen, D. Tianfu, and W. Gang. “ROV based underwater blurred image restoration”. In: *Journal of Ocean University of Qingdao* 2.1 (2003), pp. 85–88.

Own Publications

- [Far+14] Fahime Farhadifard, Elham Abar, Mahmoud Nazzal, and Huseyin Ozkaraman. “Single image super resolution based on sparse representation via directionally structured dictionaries”. In: *22nd Conference on Signal Processing and Communications Applications Conference (SIU)*. 2014, pp. 1718–1721.
- [Far15] F. Farhadifard. “Underwater Image Restoration: Effect of Different Dictionaries”. In: *International Summer School on Visual Computing*. 2015, pp. 81–90.
- [FR16] F. Farhadifard and M. Radolko. “Adaptive UW Image Deblurring via Sparse Representation”. In: *Eurographics*. 2016.
- [FRL17a] F. Farhadifard, M. Radolko, and U. Freiherr von Lukas. “Marine-Snow Detection and Removal: Underwater Image Restoration using Background Modeling”. In: *25th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision (WSCG)* (2017).
- [FRL17b] F. Farhadifard, M. Radolko, and U. Freiherr von Lukas. “Single Image Marine Snow Removal based on a Supervised Median Filtering Scheme”. In: *12th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP)*. Vol. 4. 2017, pp. 280–287.
- [FZL15] F. Farhadifard, Z. Zhou, and U. F. von Lukas. “Learning-based underwater image enhancement with adaptive color mapping”. In: *9th International Symposium on Image and Signal Processing and Analysis (ISPA)*. 2015, pp. 48–53.
- [Rad+15] M. Radolko, F. Farhadifard, E. Gutzeit, and U. Freiherr von Lukas. “Real time video segmentation optimization with a modified Normalized Cut”. In: *9th International Symposium on Image and Signal Processing and Analysis (ISPA)*. 2015, pp. 31–36.
- [Rad+16] M. Radolko, F. Farhadifard, E. Gutzeit, and U. Freiherr von Lukas. “Dataset on Underwater Change Detection”. In: *OCEANS, MONTEREY*. 2016, pp. 1–8.
- [RF16] M. Radolko and F. Farhadifard. “Using trajectories derived by dense optical flows as a spatial component in background subtraction”. In: *24th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision (WSCG)*. 2016, pp. 1–8.
- [RFL17a] M. Radolko, F. Farhadifard, and U. Freiherr von Lukas. “Background Modeling: Dealing with Pan, Tilt or Zoom in Videos”. In: *25th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision (WSCG)* (2017).

Own Publications

- [RFL17b] M. Radolko, F. Farhadifard, and U. Freiherr von Lukas. “Change Detection in Crowded Underwater Scenes - Via an Extended Gaussian Switch Model Combined with a Flux Tensor Pre-segmentation”. In: *12th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP)*. Vol. 4. 2017, pp. 405–415.

Thesis Statements

- Digital imaging has been established in a broad area of challenging topics, such as surveillance tasks, industrial quality assurance, inspection applications and exploration. This means that, digital images are encountered almost everywhere, thus, the ability to process and analyze them is of major importance.
- Despite the technological advances in the equipment, the quality of underwater images and videos remains inferior to images acquired in air. Essentially, this is because of the limitations imposed by the inherent properties of the water medium, which lead to color cast, haze, blur, low contrast and noise.
- Due to limitations in technology, live streaming of data transmitted from a camera under water using its highest resolution is still not feasible. Thus, in such situations, usually, a compressed and lower resolution version of the data is transferred.
- Besides additive noises such as sensor noise, the quality of underwater images is reduced by the presence of suspended particles in water. Such particles represent an unwanted signals, which not only decrease the scene visibility, but also is disruptive for advanced computer vision tasks, such as segmentation.
- Among the underwater image degradations, color cast and haze are addressed often. This results in a vast amount of literature on these two topics. In contrast, significant blur, limited resolution and also problems caused by the presence of marine particles are not well-researched.
- To improve the image resolution and sharpen the image features, an approach for single image super-resolution and deblurring is proposed.
 - Single image super-resolution is an ill-posed problem that requires effective regularization. Super-resolution using learned dictionaries offers a powerful framework for efficiently solving this problem by exploiting the sparsity as the regularizer.
 - Blur under water (caused by forward scattering) is a function of camera-object distance. This means that, it is spatially variant, which results in different levels of image feature blurriness. Thus it is necessary to have an adaptive approach.
 - To deal with single image super-resolution and deblurring in one unique solution, it is necessary to employ multiple coupled dictionaries. Thus, while using a powerful super-resolution approach, the spatially varying blur is also addressed.
 - Employing multiple dictionaries requires a model selection in both training and reconstruction stages. Such a classifier should have a scale-invariance property, so that the link between the low and high-resolution feature spaces is properly established.

Thesis Statements

- The presence of floating marine particles together with backscatter, has the greatest degradation factor. Thus, special approaches addressing detection and removal of such particles are proposed.
 - In order to reconstruct the scene covered by the particles, one can use the local information. However, this may lead to a poor performance of the restoration whenever the local information is not sufficient.
 - Having a video sequence, one can learn static features of the data by using the temporal information. Using such features, some statistical priors of the scene can be learned, which are then used to restore the scene.

Selbstständigkeitserklärung

Ich erkläre, dass ich die eingereichte Dissertation selbstständig und ohne fremde Hilfe verfasst, andere als die von mir angegebenen Quellen und Hilfsmittel nicht benutzt und die den benutzten Werken wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

Ich versichere weiterhin, dass ich bisher weder die vorliegende Dissertation noch Teile von ihr als Prüfungsarbeit oder zum Zweck der Promotion eingereicht bzw. verwendet habe.

Rostock, 27. October 2017

Fahimeh Farhadifard